

Draft Final Report

*Improvement of Procedures for
Evaluating Photochemical Models*

Prepared for:

*California Air Resources Board
Sacramento, California*

March 1990

RADIAN
CORPORATION



20th★★★★
Anniversary

RADIAN
CORPORATION

264-069-05-00

DCN: 90-264-069-05-01

10395 Old Placerville Road
Sacramento, CA 95827
(916) 362-5332

**IMPROVEMENT OF PROCEDURES FOR EVALUATING
PHOTOCHEMICAL MODELS**

**Draft Final Report
Contract No. A832-103**

Prepared for:

Research Division
California Air Resources Board
1102 Q Street
P.O. Box 2815
Sacramento, CA 95812

Submitted by:

Radian Corporation
10395 Old Placerville Road
Sacramento, CA 95827

Prepared by:

T. W. Tesche
Principal Investigator
Alpine Geophysics
Placerville, CA 95667

Panos Georgopoulos
John H. Seinfeld
Glen Cass
California Institute of
Technology
Pasadena, CA 91125

Fred L. Lurmann
Sonoma Technology, Inc.
Santa Rosa, CA 95403

Philip M. Roth
San Rafael, CA 94960

March 7, 1990

ACKNOWLEDGMENTS

The authors acknowledge Mr. Bart E. Croes of the California Air Resources Board (ARB), Research Division, for providing guidance on the overall direction of the study, and Dr. Kit K. Wagner of the ARB Technical Support Division for several helpful discussions and insights. This report was submitted in fulfillment of Contract No. A832-103, "Improvement of Procedures for Evaluating Photochemical Models," by Radian Corporation under sponsorship of the California Air Resources Board. The statements and conclusions in this report are those of the contractor and not necessarily those of the California Air Resources Board. The mention of commercial products, their source, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products.



TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
ACKNOWLEDGEMENTS	ii
LIST OF TABLES	vi
LIST OF FIGURES	vii
1.0 INTRODUCTION	1-1
1.1 Overview	1-1
1.2 Review of Previous Work	1-4
1.3 Study Objectives	1-6
1.4 Report Structure	1-7
1.5 Definition of Terms	1-8
2.0 OPERATIONAL MODEL PERFORMANCE EVALUATION PROCEDURES	2-1
2.1 Numerical Procedures	2-1
2.1.1 Accuracy	2-3
2.1.2 Mean Bias	2-8
2.1.3 Variance	2-9
2.1.4 Gross Error	2-10
2.1.5 Use of the Numerical Measures	2-10
2.2 Graphical Procedures	2-12
2.2.1 Accuracy Plot	2-13
2.2.2 Time Series Plots	2-16
2.2.3 Spatial Time Series Plots	2-16
2.2.4 Ground Level Isopleths	2-20
2.2.5 Scatter Plot of Predictions and Observations	2-20
2.2.6 Scatter Plot of Residuals and Observations	2-24
2.2.7 Bias Stratified by Concentration	2-27
2.2.8 Gross Error Stratified by Concentration	2-27
2.2.9 Bias Stratified by Time	2-30
2.2.10 Gross Error Stratified by Time	2-30

TABLE OF CONTENTS (Continued)

<u>Section</u>	<u>Page</u>
2.3 Diagnostic Simulations	2-33
2.3.1 Zero Emissions	2-33
2.3.2 Zero Initial Conditions	2-36
2.3.3 Zero Boundary Conditions	2-38
2.3.4 Zero Surface Deposition	2-38
2.3.5 Increased Height Variations	2-39
2.3.6 Reduced Wind Speeds	2-41
2.4 Guidelines for Assessing Simulation Results	2-41
3.0 DIAGNOSTIC MODEL PERFORMANCE EVALUATION PROCEDURES	3-1
3.1 Sensitivity-Uncertainty Analysis	3-1
3.1.1 Sensitivity-Uncertainty Analysis	3-4
3.1.2 Characterization of Uncertainties and Comparisons of Uncertain Predictions and Observations	3-6
3.2 Analysis of Residuals	3-7
3.3 Multi-Species Comparisons	3-10
3.3.1 Selection and Organization of Predicted/Observed Data	3-11
3.3.2 Performance Measures and Statistical Tests	3-12
3.3.3 Graphical Procedures	3-13
3.3.4 Measurements and Measurement Evaluation	3-13
3.3.5 Model Modification to Facilitate Testing	3-15
3.4 Mass Fluxes and Budgets	3-16
4.0 RESEARCH NEEDS IN MODEL PERFORMANCE EVALUATION	4-1
4.1 Testing the Adequacy of Model Response to Emission Changes	4-1
4.2 Mechanistic Model Evaluations	4-4
4.2.1 Emissions Model Testing	4-4
4.2.2 Meteorological Model Testing	4-7
4.2.3 Chemistry Model Testing	4-17
4.2.4 Deposition Model Testing	4-25

TABLE OF CONTENTS (Continued)

<u>Section</u>	<u>Page</u>
4.3 Non-Traditional Approaches to the Incommensurability Problem	4-30
4.3.1 Sub-grid Scale Modeling	4-31
4.3.2 Finer Grid Resolution	4-32
4.3.3 Remote Measurements	4-32
4.3.4 Multiple Measurements Within a Cell	4-33
4.4 Stressing a Model	4-33
4.5 Use of Simplified Models	4-38
4.5.1 Photochemical Box Models	4-38
4.5.2 Photochemical Trajectory Models	4-39
4.5.3 Suggested Uses of Simplified Photochemical Models	4-41
5.0 SUMMARY AND RECOMMENDATIONS	5-1
5.1 Summary	5-1
5.2 Recommendations	5-2
5.2.1 Operational Evaluation Procedures	5-2
5.2.2 Optional by Strongly Recommended Evaluation Procedures	5-11
5.2.3 Longer Term Needs in Performance Evaluation	5-13
6.0 BIBLIOGRAPHY	6-1

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Typical Questions That Arise in Photochemical Model Applications in Support of Regulatory Decision-Making	1-3
Maximum Concentrations for Base Case and Moderate Emissions Reduction Simulations	2-37
Recommended Numerical Measures for Evaluating Diagnostic and Prognostic Meteorological Models	4-13

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
2-1 Recommended Operational Performance Evaluation Measures and Procedures	2-2
2-2 Different Measures of Peak Prediction Accuracy	2-14
2-3 Accuracy of Station Peak Predictions	2-15
2-4 Ozone Time Series Plot at Thousand Oaks for 6-7 September 1984 Base Case in Ventura-Santa Barbara	2-17
2-5 Spatial Ozone Time Series Plot at Piru for 25-26 September 1980 Base Case in Ventura-Santa Barbara	2-19
2-6 Ground Level Hourly Average Ozone Concentration Isopleth on 17 September 1984 Between the Hours of 1500 and 1600 PST (pphm) . .	2-21
2-7 Correlation of Hourly Average Predictions and Observations at All Monitoring Stations	2-22
2-8 Correlation of Daily Maximum Predictions and Observations at All Monitoring Stations	2-23
2-9 Distribution of Hourly Concentration Residuals as a Function of Observed Concentration	2-25
2-10 Distribution of Maximum Concentration Residuals as a Function of Observed Concentration	2-26
2-11 Bias as a Function of Observed Concentration Level	2-28

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
2-12 Gross Error as a Function of Observed Concentration Level	2-29
2-13 Error as Function of Simulation Time	2-31
2-14 Bias as a Function of Simulation Time	2-32
2-15 Deficit-Enhancement Plot of the Effects of Moderate Emissions Reductions on Ground Level Ozone Concentrations - 1600 PST on 6 September 1984	2-34
2-16 Time Series Plot of Base Case and Zero Emissions Diagnostic Run at the Pasadena Monitoring Station	2-35
2-17 Time-Space Uncertainty Envelopes of Ozone Prediction and Measurement Uncertainties on 17 September 1984 at the Simi Monitoring Station	2-40
4-1 North-South Slice of Modeling Region Depicting Terrain, Temper- ature Structure, and Mixing Height. 1600 PST, 25 September 1980 . . .	4-16
4-2 Predicted Surface Winds for 11 September 1980. 0600-0700 LST . . .	4-18
4-3 Comparison of Estimated and Observed Hourly Wind Values at the Ventura Station in the SCCAB - 25 September 1980	4-19
4-4 Residual Distributions of Predicted and Observed Wind Speeds . . .	4-20
4-5 Statistical Comparison of Observed and Modeled Surface Winds With Time on 8 August 1984	4-21

LIST OF FIGURES (Continued)

<u>Figure</u>		<u>Page</u>
4-6	Hierarchy of Species for Mechanism Testing	4-24
4-7	Normalized Bias (%) in Maximum Ozone in Simple and Complex Organic Mixture Runs	4-26



1.0 INTRODUCTION

1.1 Overview

Urban ozone modeling techniques have matured significantly since the original development work in the South Coast Air Basin two decades ago. Despite much success, important problem areas remain. Possibly the most pressing is how to determine whether a photochemical model performs well enough for use in regulatory decision-making. There are several facets to this problem.

Many decision makers have limited confidence in the use of models in the planning process, noting that ozone concentrations haven't declined over the last several years as much as models predicted they would (API, 1989). Many counties and metropolitan areas identified by the United States Environmental Protection Agency (EPA) in 1979 as non-attainment should have reached attainment by 31 December, 1987. But only 39 percent of these areas met the ozone standard. Is this a result of inadequate models and data bases? No doubt, part of the answer is that many planned emissions reduction measures were not implemented because control technology was not fully developed or was considered too costly. Some control measures were not as effective as originally thought; others were not enforced. Nevertheless, many of the State Implementation Plans (SIPs) understated the amount or type (e.g., oxides of nitrogen or reactive organic gases) of control needed as a result of inadequate data and models.

Sometimes the desired reductions in ozone concentrations predicted by models are of the same magnitude as the prediction error. In such a case, can the model be used to predict ozone attainment for such a metropolitan region? Can photochemical models be used to assess the impact of single sources or the effects of emissions control strategies involving only small emissions changes over the urban area?

Often only one or two episodes are modeled and their representativeness of the broader range of conditions is questioned. Thus, how many episodes must be studied to test rigorously a proposed control program? Should the episodes consist of severe ozone-producing conditions or should a few less severe episodes

be studied? The Environmental Protection Agency (Schere and Shreffler, 1982) suggest that the most stringent emissions control requirements may, in some cases, result from less severe episodes. What, then, is the optimal collection of episodes to model?

Traditional photochemical model performance evaluations do not provide sufficient information to decision-makers about the suitability of a model for use in regulatory applications (Dennis and Downton, 1984; Seinfeld, 1988a,b; Tesche and Roth, 1988). Typical questions that arise are listed in Table 1-1.

Present photochemical model evaluation methodologies are largely "operational." They are not specifically designed to reveal flaws in the model, the data base, or the procedures used to exercise the model. Instead, they are aimed at quantifying the correspondence between predicted and observed ground-level concentrations. While current operational evaluation procedures may indicate to the decision-maker how well a model performed in an overall sense relative to past, similar applications, little direct information is provided about the model's suitability for predicting the effects of emissions reductions.

To date, there have been no standardized procedures for conducting model performance evaluations. Also lacking are agreed-upon procedures for rejecting simulation results. Past ozone modeling protocols have accepted photochemical model results as "adequate" if the bias and error statistics were "comparable" with those of previous, similar studies, but little attention was given to whether the previous simulations adequately met policy-making needs. Furthermore, previous model evaluations have not rigorously pursued sufficient diagnostic tests (e.g., mass budgets and tests of precursor species) or mechanistic evaluations (e.g., deposition and meteorological module testing) to confirm that the ozone results are "correct for the right reason."

New procedures are needed to deal with these issues. The performance of photochemical models, judged by commonly-reported statistical procedures, appears to have reached a plateau (Seinfeld, 1988a,b). Current photochemical grid models reproduce hourly averaged ozone concentrations to within 30-35

TABLE 1-1. TYPICAL QUESTIONS THAT ARISE IN PHOTOCHEMICAL MODEL APPLICATIONS IN SUPPORT OF REGULATORY DECISION-MAKING.

-
- If successfully evaluated for a base case, does the model hold for a significantly reduced level of emissions?
 - What is the probability that a desired air quality level will indeed be achieved, given the application of a deterministic model to an inherently stochastic problem?
 - Has the model been adequately "stressed" in prior testing, i.e., if it is significantly flawed, will it have failed demonstrably? Have the evaluation tests been sufficiently stringent to cause a flawed model to fail?
 - What is the risk of underestimating or overestimating the impact of a control strategy?
 - What are the component uncertainties associated with the modeling analysis and what is the aggregate uncertainty in the control strategy requirement?
 - Is an average error of, say 30 percent, in hourly averaged ozone predictions low enough?
 - Is a 10 percent underprediction bias acceptable when developing multi-billion dollar emission control programs? How does one account for this bias?
 - How much confidence can be placed in a photochemical model simulation exhibiting "good" ozone performance statistics when the precursor and product species (NO, NO₂, VOCs, PAN) are grossly and systematically underestimated?
 - Does a model evaluation whose ozone performance statistics are "comparable with other similar studies" provide any assurance that the predictions are correct for the right reason or that control strategies based upon the modeling will work?
 - Can a photochemical model be used to demonstrate attainment when the best estimate of its ozone prediction error is of the same magnitude as the ozone concentration reduction needed to meet the standard?
 - No matter how well a model performs in a base case, can it be trusted to predict the future consequences of a strategy implemented now?
-

percent and the peak one-hour prediction is often reproduced to within 15-20 percent (Tesche, 1988). In spite of the wealth of routine statistics on model bias, error, and correlation gained from literally hundreds of simulations, many decision-makers have limited confidence in using models and their predictions (Roth et al., 1989). To assist decision-makers further, model performance evaluation methodologies need to be expanded to include diagnostic, comparative, and mechanistic tests in addition to those currently being performed.

1.2 Review of Previous Work

Nearly a decade ago, the U. S. Environmental Protection Agency commissioned several studies to develop procedures for evaluating photochemical grid models. Among the relevant studies are those of Hayes (1979); Hillyer, Reynolds, and Roth (1979); Bencala and Seinfeld (1979); Dennis and Downton (1982); Schere and Shreffler (1982); Moore et al., (1983), and Tesche et al., (1983). These and other investigators produced a set of quantitative and qualitative procedures for evaluating photochemical model performance.

In a more general modeling context, the highly-regarded American Meteorological Society (AMS) workshop on model evaluation developed a natural hierarchy of performance measures and accompanying statistical methods (Fox, 1981). Eight types of graphical display methods were identified in addition to the numerical procedures. The AMS procedures may be grouped as follows:

- Pearson moment correlation coefficient;
- Time lagged cross-correlation function;
- Goodness of fit tests for cumulative frequency distributions;
- Bias as measured by the average of the residuals;
- Noise as measured by the variance of the residuals;
- Gross error as measured by the root mean square error;

- Vector differences in locations between the highest observed and predicted concentrations; and
- Time differences between the occurrence of the highest observed and predicted concentrations.

Because the AMS workshop was not restricted to photochemical grid models, it is not surprising that several of the recommended numerical and graphical procedures (e.g., correlation coefficients, vector difference plots) have not been widely used in ozone model evaluations. Several performance measures not identified by the AMS workshop or the earlier EPA studies have also been proposed, including ratios of prediction and observation; the percentage of predictions lying within N-percent of the observation; least squares regression statistics based on prediction and observation; largest positive and range of errors; coefficients of variation of the residuals where the standard deviation is normalized by the mean observation or prediction; distance distributions; and time series plots using "best prediction" values for comparisons with observations.

Despite over a decade of accumulated experience in photochemical model evaluation and refinement, there has been no focused attempt by regulatory agencies, model developers, or others to derive a consistent set of evaluation procedures from among the techniques in common usage, notwithstanding the widely recognized need for such methodology. The major problem areas are:

- Lack of consistency among performance evaluations conducted by different groups; for example, various definitions of ozone "cutoff levels" or time periods over which performance is quantified;
- Use of alternative definitions of certain key performance measures such as bias, error, or peak prediction accuracy;
- Ambiguous meanings of some correlation measures such as the spatial correlation coefficient; and

- Inability of most measures to reveal clearly the existence of model flaws, data base deficiencies, or both.

This study attempts to fulfill this need for a consistent set of evaluation procedures and to provide a basis for addressing current issues including those previously identified. Results of this study may be incorporated into the ARB's guidelines on photochemical modeling in California (ARB, 1989).

1.3 Study Objectives

Our goal in this study is to establish the basis for consistent photochemical grid model performance evaluations in the near term and to provide a framework for performance evaluation research over the longer term. The specific study objectives include:

- Develop improved evaluation procedures for photochemical grid models;
- Develop evaluation procedures supporting regulatory decision-making concerning ozone control programs;
- Develop new procedures consistent with improved aerometric data bases such as the Southern California Air Quality Study (SCAQS); and
- Exemplify the use of the recommended procedures and performance measures.

We have attempted to develop a sound framework for photochemical model performance testing, incorporating several procedures and measures that have proven useful in the past. These existing procedures are supplemented with new methods where appropriate. Specific suggestions are made on how additional, newer evaluation procedures and measures may be developed as measurement data bases become more robust and as the discipline of model evaluation matures. The

evaluation methodology suggested here will require updating as improved information and procedures become available.

1.4 Report Structure

Section 2.0 of this report presents the statistical and graphical performance evaluation procedures that we recommend for routine use in photochemical modeling studies in California. Included is a set of diagnostic simulations that should accompany the performance evaluations. Tabular and graphical examples of many of these measures are provided to suggest ways in which the results might be portrayed.

Section 3.0 discusses diagnostic model evaluation methods. Strongly recommended for routine performance evaluations, these procedures are intended to develop greater insight into the strengths and weaknesses of a particular model, data base and photochemical simulation than is afforded by the routine procedures suggested in Section 2.0. In difficult modeling situations such as the South Coast Air Basin and the San Francisco Bay Area, many of the diagnostic procedures presented in Section 3.0 are an essential component of the overall performance evaluation process.

Recognizing that photochemical modeling is a dynamic process, Section 4.0 addresses emergent and longer term performance evaluation issues and research needs. In some cases, these needs have been identified for some time; only recently have data bases and numerical modeling techniques become available to address them. In time, some of the research needs discussed in Section 4.0 will be addressed successfully, leading to an improved set of model and data base testing procedures.

Section 5.0 contains the Summary and Recommendations. We provide an expanded bibliography on photochemical modeling in Section 6.0 in lieu of a reference section to aid individuals in locating potentially relevant photochemical modeling studies.

1.5 Definition of Terms

Several different terms are used in the literature to refer to the mixture of organic compounds that are emitted from natural and man-made sources and that react in the atmosphere to form ozone. Terms such as non-methane hydrocarbons (NMHC), reactive hydrocarbons (RHC), volatile organic compounds (VOC's), reactive organic gases (ROG), and non-methane organic compounds (NMOC) are occasionally used interchangeably to describe the mixture of compounds that are emitted from sources, measured in the atmosphere with instruments, or are simulated with mathematical models and chemical mechanisms. In this report we use ROG to refer to emissions of reactive organics (expressed in units of mass per time) and VOC to signify the volumetric concentration of species measured in the atmosphere or simulated by photochemical models.

Another set of definitions relates directly to performance evaluation. Following are the definitions of specific terms that are used repeatedly throughout this report:

Accuracy: The degree, usually expressed in percent, to which a one hour average prediction deviates from the corresponding hourly-averaged measurement. The predicted value may be the highest one hour concentration during the episode, the peak value at a given monitoring station, or some other value.

Bias: The first moment of the distribution of concentration residuals. Bias provides a measure of the overall tendency of the model to over- or underestimate the observed concentration fields.

Comparative Evaluation: An intercomparison between two different photochemical models or different modules (e.g. wind field generators, mixing height algorithms, chemical kinetic mechanisms) using the observed fields as the standard against which the intercomparisons are made.

Cutoff Level: A predetermined observed concentration level below which pairs of predicted and observed values are not considered in calculating various statistical relative difference measures. The cutoff level is used to avoid

unduly influencing the calculation of relative difference measures by allowing near zero observed concentrations to appear in the denominator of certain performance measures. Although inherently arbitrary in specification, previous studies have generally set cutoff levels of 6 pphm for ozone, 2 pphm for NO and NO₂, and 1 pphm for VOC concentrations. These values are suggested here.

Diagnostic Evaluation: An assessment of a model's ability, when functioning as a whole, to predict specific details or processes occurring during a photochemical episode (e.g., dispersion patterns, deposition rates). The events and tests are specifically chosen to challenge the science in the model. Specific focus is on detailed examination of how well individual components of the model (i.e., modules) simulate actual atmospheric processes.

Error: The mean absolute deviation among the hourly-averaged predicted concentrations and the corresponding measurements.

Mechanistic Evaluation: An assessment of the ability of an individual process modules' ability to reproduce the observed salient features of the process. When applied to all process modules that constitute the full model, it represents a test of the correctness of the underlying science.

Model Evaluation: Part of the continuing process of model development, data collection, model testing, diagnostic analysis refinement, and retesting. While a single evaluation cannot validate a model, it can raise serious doubts about the model's adequacy. Model evaluation represents testing of a model's ability to predict accurately observed measures of air quality over a range of air quality and meteorological conditions.

Model Validation: A potential outcome of the process of model evaluation and verifications. The "validity" of a model refers to how well model predictions would agree with the appropriate observations given a perfect specification of model inputs. Validity refers to the inherent quality of the model formulation. Model validation is the establishment of a conclusion by detailed and copious evidence that leads to formal recognition. This may include several model performance evaluations.

Model Verification: A successful or positive outcome of the model evaluation process. Verification is proof of the accuracy, reality, or truth of the model.

Operational Evaluation: An assessment of a model's ability to predict the correct answer whether or not the process descriptions in the model are accurate. It is an examination of how well a model reproduces observed concentration fields in time and space consistent with the needs of policy analysis or regulatory decision making.

Performance Evaluation: The process of testing a model's ability to predict accurately observed measures of air quality over a range of meteorological and air quality conditions. Testing the predictive capabilities of supporting models or modules, such as chemical mechanisms and wind field models, is an important component of performance evaluation.

Precision: The second moment of the distribution of concentration residuals.

Residual: The difference, in units of concentration, between a predicted value and the value observed at a monitor at the same location and time. The concentration residual forms the basis for all difference measures used in this study.

Sensitivity Analysis: The process of studying (through numerical simulation) the influence on model predictions due to variations in one or more model inputs, including emissions, meteorological, or air quality parameters. One component of sensitivity analysis is the variation of inputs over their range of uncertainty in order to develop estimates of the range in predicted concentrations associated with to these input uncertainties.

Variance: The second moment of the distribution of concentration residuals. This statistic is commonly used as a measure of the precision of a simulation.

Mathematical definitions and procedures for estimating and interpreting these quantities are addressed in the following three sections.

2.0 OPERATIONAL MODEL PERFORMANCE EVALUATION PROCEDURES

Before a photochemical model may be applied to simulate the effects of emissions controls, it is appropriate to assess whether the model adequately reproduces the chemical and physical processes governing urban-scale ozone concentrations. Referred to as model performance evaluation, this assessment involves the compilation of emissions, meteorological, air quality and chemical data, drawn from an historical ozone episode, and the use of this information in the exercise of the model. Through various statistical and graphical means, the model predictions and air quality observations are compared to determine the "fidelity" of the simulation. The first set of simulation results seldom provides a level of correspondence with the observations that is judged to be adequate. Consequently, the model inputs and outputs are reexamined to ascertain the cause(s) of poor performance. Where modifications to the inputs or model algorithms are justified, these changes are made and the photochemical model is re-run. This process is repeated until adequate performance is achieved or until no further modifications to the model inputs can be justified.

Specific numerical and graphical procedures are recommended in this section for operational evaluation of the performance of grid-based photochemical dispersion models. The methods suggested include the calculation of certain peak prediction accuracy indices, various statistics based on concentration residuals, and time series of predicted and observed hourly concentrations. Graphical procedures are suggested to complement the numerical measures, providing additional insight into model performance. Finally, a minimum set of six diagnostic simulations are recommended as part of operational performance evaluations. These procedures are listed in Figure 2-1.

2.1 Numerical Procedures

Ten numerical measures may be used to typify a model's overall performance in a photochemical simulation. Several of these measures, involving specific comparisons between hourly predictions and observations, have been used

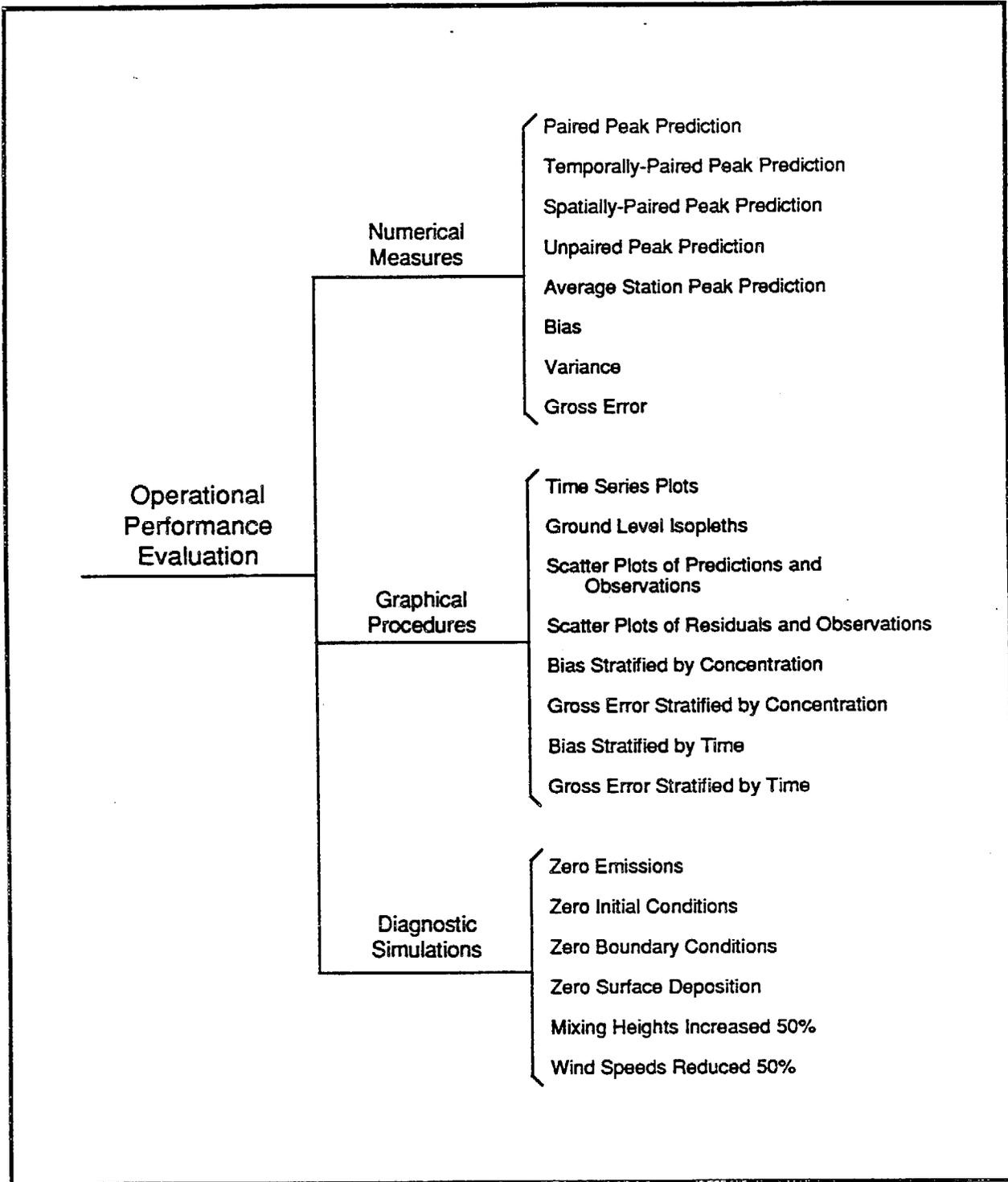


Figure 2-1. Recommended Operational Performance Evaluation Measures and Procedures.

extensively in past performance evaluations, though not always consistently. As shown in Figure 2-1, the measures involve the accuracy of peak one-hour average prediction, bias, variance, and error. These measures may be applied to any primary or secondary pollutant for which adequate monitoring data are available. ~~For the present, however, their use is recommended primarily for ozone and NO_x concentrations.~~

2.1.1 Accuracy

The accuracy of a photochemical model's prediction of peak one-hour average concentrations may be estimated in five ways. Several accuracy measures are proposed because there are different ways of defining "peak" concentrations. These five measures provide useful tests of a model's performance, particularly from a regulatory perspective, but they convey only a portion of the information needed to assess the overall adequacy of a model simulation. The real value of the five measures will likely be determined only after their usage in a number of photochemical model evaluation studies by various groups of investigators.

Paired Peak Prediction Accuracy. The paired peak prediction accuracy, A_{ts} , is given by:

$$A_{ts} = \frac{c_p(\hat{x}, \hat{t}) - c_o(\hat{x}, \hat{t})}{c_o(\hat{x}, \hat{t})} \times 100\% \quad (2-1)$$

where c_p is the predicted one-hour pollutant concentration, c_o is the observed hourly averaged concentration, \hat{x} refers to the peak monitoring station location, \hat{t} is the time of the peak observation. The caret, $\hat{}$, denotes the time or location of the maximum observed concentration. A_{ts} quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station, $c_o(\hat{x}, \hat{t})$, and the predicted concentration at the same location, \hat{x} , and at the same time, \hat{t} . Predictions and observations are thus "paired in time and space." The paired peak prediction accuracy is a stringent model evaluation measure. It quantifies the model's ability to reproduce, at

the same time and location, the highest observed concentration during each day of the episode.

There are at least three different definitions of the "predicted" concentration that may be used in calculating the accuracy measures and other statistics. These three procedures are discussed later in Section 2.2. ~~Unless otherwise stated, the predicted concentration used in all comparisons with observations is the four-cell weighted average derived through simple bilinear interpolation~~ (Press et al., 1986) of the predictions in the nearest four grid cells. Obviously, the monitoring station will fall within one of the four grid cells. Other, higher order interpolation schemes might be used (e.g. bicubic interpolation, inverse distance square weighting, bicubic spline) but these more sophisticated methods are not justified in light of the models minimum theoretical spatial resolution of approximately two grid cells.

A_t is very sensitive to spatial and temporal misalignments between the predicted and observed concentration fields. These space and time offsets may arise from spatial displacements in the transport fields resulting from biases in wind speed and direction, problems with the "timing" of photochemical oxidation and removal processes, or subgrid-scale phenomena (e.g., ozone titration by local NO_x emission sources) that are not resolvable by the model.

Temporally-Paired Peak Prediction Accuracy. The temporally-paired peak prediction accuracy, A_t , is given by:

$$A_t = \frac{c_o(\hat{x}, \hat{t}) - c_p(\hat{x}, \hat{t})}{c_o(\hat{x}, \hat{t})} \times 100\% \quad (2-2)$$

A_t quantifies the discrepancy between the highest concentration measurement at a monitoring station and ~~the highest prediction at the same or any other monitoring station that is located within a distance of 25 km. This distance is suggested as a guideline.~~ In heavily populated air basins in California such as the South Coast, South Central Coast, and Bay Area, often there are three or more monitoring stations within a 25 km radius.

~~The temporally paired peak prediction accuracy examines the model's ability to reproduce the highest observed concentration in the same subregion at the correct hour.~~ Coupled with other performance measures, it provides some insight into the reasonableness of the model's chemical transformation processes that lead to concentration maximum.

Relaxation of the spatial-pairing requirement (compared with A_t) may be appropriate if the spatial offsets between predicted and observed concentrations are the ~~result of microscale wind patterns, locally complex terrain, peculiarities in the siting of individual monitoring stations, or other factors.~~ In large air basins or where different subregional source-receptor flow patterns exist, the monitor where the highest prediction occurs may not correspond to the location where the highest observation was recorded. For example, in the South Coast Air Basin, it is inappropriate to calculate A_t based on a peak measurement at Riverside (the eastern basin) and a peak prediction at Reseda (in the San Fernando Valley). ~~The monitor where peak prediction is used in Equation 2-2 should experience similar source-receptor phenomena as the location of the observed maximum, i.e., it should be in the same subregion.~~ As noted, a maximum distance of 25 km is suggested as an estimate of subregion size, although this criterion may need to be revised depending upon the air basin and its specific monitoring network.

Spatially-Paired Peak Prediction Accuracy. The spatially-paired peak prediction accuracy, A_s , is given by:

$$A_s = \frac{c_p(\hat{x}, t) - c_o(\hat{x}, \hat{t})}{c_o(\hat{x}, \hat{t})} \times 100\% \quad (2-3)$$

A_s quantifies the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station and the highest predicted concentration at the same monitor, ~~within a specified time interval centered on the peak hour.~~ A ~~minimum temporal offset of three hours is recommended~~ based on review of typical predicted and observed ozone time series profiles in several California air basins. ~~If the model is unable to replicate the time of the observed~~

~~peak to within three hours, this probably indicates a major problem with the model or input files.~~ When interpreted along with other performance quantities, this measure may provide some insight into the reasonableness of the simulated transport processes leading to the concentration maximum.

Relaxation of the temporal-pairing requirement (compared with A_{ts}) may be appropriate if the causes of temporal offsets between predicted and observed concentrations result from circumstances that do not compromise the model's utility in control strategy evaluation. For example, situations involving extended transport times across large air basins, multiple-day episodes with significant pollutant recirculation, or episode days with considerable air mass stagnation prior to the time of the ozone peak may be difficult for a model to transport the peak ozone cloud correctly. ~~This difficulty in matching the exact time of the observed peak is regularly encountered in photochemical model evaluation studies.~~ It is not uncommon to find a one to two hour offset between predicted and observed ozone maximum.

Unpaired Peak Prediction Accuracy. The unpaired peak prediction accuracy, A_u , is given by:

$$A_u = \frac{c_p(x, t) - c_o(\hat{x}, \hat{t})}{c_o(\hat{x}, \hat{t})} \times 100\% \quad (2-4)$$

A_u quantifies the difference between the magnitude of the peak one-hour average concentration and the highest predicted value in the modeling domain. The unpaired peak prediction accuracy tests the model's ability to reproduce the highest observed concentration anywhere in the region. This is the least stringent of the four peak prediction accuracy measures introduced thus far. This measure quickly identifies situations where the model produces maximum ozone concentrations in the air basin that significantly exceed the highest observed values within the network.

Average Station Peak Prediction Accuracy. The average station peak prediction accuracy, \bar{A} , is given by:

$$\bar{A} = \frac{1}{N} \sum_{i=1}^N |A_{s,i}| \quad (2-5)$$

where:

$$A_{s,i} = \frac{c_p(\hat{x}_i, \hat{t}) - c_o(\hat{x}_i, \hat{t})}{c_o(\hat{x}_i, \hat{t})} \times 100\% \quad (2-6)$$

\bar{A} is the mean of the spatially-paired peak prediction accuracies ~~averaged over~~
~~all monitoring station locations.~~ \bar{A} is calculated from Equation 2-5 by first
determining the spatially-paired peak prediction accuracy at each monitoring
station, $A_{s,i}$, based upon Equation 2-6. ~~Only those monitoring stations whose peak~~
~~observed concentrations equal or exceed the cutoff level are included in the~~
~~analysis.~~

The average station peak prediction accuracy is simply the mean of the
absolute value of the $A_{s,i}$ scores (Equation 2-6). ~~It is recommended using the~~
~~spatially-paired peak prediction accuracy measure to compute the average value,~~
instead of the temporally-paired metric, in part because the former has been used
far more often in evaluation studies. As noted earlier, the temporal offset
between predicted and observed maxima at any monitoring station should not exceed
three hours.

The average station peak prediction accuracy describes how well the
maximum concentrations at individual stations throughout the monitoring network
are reproduced. ~~This measure is useful in comparing different model simulations~~
~~and for assessing the peak prediction performance over the whole modeling domain.~~
By itself, \bar{A} does not reveal the adequacy of a simulation for regulatory purposes
because the averaging process ~~tends to degrade the model's performance at the~~
~~highest monitoring station.~~

2.1.2 Mean Bias

The mean bias (sometimes called the mean bias error) is calculated as a residual quantity and as one that is normalized by the observed concentrations. The non-normalized bias, D, is given by:

$$D = \frac{1}{N} \sum_{i=1}^N (c_p(x_i, t) - c_o(x_i, t)) \quad (2-7)$$

where N equals the number of hourly prediction-observation pairs drawn from all monitoring stations for the particular day. The normalized bias is:

$$D^* = \frac{1}{N} \sum_{i=1}^N \frac{(c_p(x_i, t) - c_o(x_i, t))}{c_o(x_i, t)} \quad (2-8)$$

The bias is based on the average signed deviation of the concentration residuals. It indicates the degree to which predicted one-hour concentrations are over- or underestimated. Based on the ensemble of prediction-observation pairs, this measure reveals the presence of systematic deviation from observed concentrations. Both non-normalized (Equation 2-7) and normalized estimates (Equation 2-8) of bias are recommended. The former provides an indication of the overall tendency of the model to over- or underestimate in units of concentration. Model prediction bias may thus be compared directly with the ozone design value and ambient air quality standards. The normalized bias is very useful because it facilitates direct comparisons between different diagnostic simulations of a given episode and comparisons between photochemical model performance across different urban areas experiencing widely varying concentration levels.

Bias estimates are useful in identifying systematic errors in the model's temporal or spatial response. Because the bias reveals the tendency for systematic over- or underestimation, it should be zero in the ideal case. Both normalized and non-normalized (absolute) quantities should be reported, in part to facilitate comparisons with other studies where the range of concentration

levels differs. For the most part, the mean normalized bias is of greater regulatory interest so, hereafter, we refer to it as simply the bias.

Bias estimates provide little direct insight into systematic under- or overestimation problems that may exist on a subregional basis, or within specific time periods during the diurnal cycle. By itself, the bias sheds little light on the causes of the systematic under- or overprediction, and in fact, low estimates may conceal significant bias. It is possible, for example, for a model simulation to exhibit strong underprediction in one region and strong overprediction in another region. On average, the regionwide bias estimate could be zero, yet significant local biases would still remain. Small or zero bias thus does not guarantee good model performance since large gross errors (and variance) can accompany a zero first moment. Obviously, bias must be interpreted carefully along with the other residual measures.

2.1.3 Variance

The variance of the residual distribution, S_d^2 , is given by:

$$S_d^2 = \frac{1}{N-1} \sum_{i=1}^N (d_i - D)^2 \quad (2-9)$$

where the concentration residual is defined as:

$$d_i = c_p(x_i, t) - c_o(x_i, t) \quad (2-10)$$

and D is the first moment, i.e., the mean non-normalized bias. This statistic describes the "dispersion" or spread of the residual distribution about the mean. The variance is calculated using all prediction-observation pairs above the cutoff level. The variance is the second moment of the residual distribution, whereas the standard deviation is simply the square root of the variance. It measures the average "spread" of the residuals, independent of any systematic bias in the predictions. No direct information is provided about subregional

errors or about large discrepancies occurring within portions of the diurnal cycle.

2.1.4 Gross Error

The mean gross error is calculated in two ways, similar to the bias. The mean absolute gross error, E_d , is given by:

$$E_d = \frac{1}{N} \sum_{i=1}^N |c_p(x_i, t) - c_o(x_i, t)| \quad (2-11)$$

The mean absolute normalized gross error is:

$$E_d^* = \frac{1}{N} \sum_{i=1}^N \frac{|(c_p(x_i, t) - c_o(x_i, t))|}{c_o(x_i, t)} \quad (2-12)$$

The gross error quantifies the mean absolute signed deviation of the concentration residuals. It indicates the average (signed) discrepancy between hourly predictions and observations. Gross error is one of the most useful single measures for comparing different model simulations. Normalized and absolute (non-normalized) estimates of the gross error should be reported, but the normalized value is generally the more useful in assessing the quality of a specific simulation. Gross error is a robust measure of overall model performance and provides a useful basis for comparison among model simulations across different air basins or ozone episodes. Compared with the variance, the gross error is somewhat less sensitive to extreme values (Fox, 1981). It provides no direct information about sub-regional errors or about large discrepancies occurring within portions of the diurnal cycle.

2.1.5 Use of the Numerical Measures

Ten numerical measures have been suggested for operational performance evaluations. Their estimation and interpretation serve two main purposes.

First, and most important, these measures provide a basis for judging whether a photochemical simulation is an adequate representation of the episode modeled. A subset of the ten numerical measures, identified below, appear well-suited for this purpose. Second, routine reporting of the full set of measures will add, in time, to the growing knowledge base for this class of atmospheric models. In particular, recent, comprehensive surveys of photochemical model evaluation, sensitivity, and uncertainty studies have revealed a lack of uniform procedures for quantifying model performance or for reporting sensitivity/uncertainty simulation results (Tesche, 1988; 1989). By adopting, consistently reporting, and archiving results from a standardized set of performance measures, the modeling community will develop a data base against which new performance evaluation studies may be compared. Not only will this data base facilitate judgements regarding the acceptability of new modeling results, but it will also provide much needed information on the causes of model failures so that these situations may be avoided in the future.

Four numerical measures appear to be most helpful in making an initial assessment of the adequacy of a photochemical simulation. These include:

- The paired peak prediction accuracy, A_{ps} ;
- The unpaired peak prediction accuracy, A_u ;
- The mean normalized bias, D^* ; and
- The mean absolute normalized gross error, E_d^* .

These measures alone are insufficient in many cases to describe adequately a model's performance. Where this occurs, the full set of operational measures, together with the graphical procedures and diagnostic simulations will provide additional, useful information. Section 2.4 addresses the difficult issue of setting model performance standards.

Confidence intervals or statistical significance tests are specifically not recommended as part of operational performance evaluation. The computation and interpretation of statistical significance for photochemical model results is fraught with a number of uncertainties and potentially stringent limitations (e.g. statistical independence, distributional assumptions), and interpretation

of results based on these assumptions can be very misleading. Preisendorfer and Barnett (1983) characterize this problem well "...because the sample is small and its parts may not always be statistically independent, one cannot confidently rely on any one of the classical continuous probability density functions such as the normal, Student-t, chi-square and F-distributions. The limited populations of real fields and associated computer-generated fields are too sparse and statistically interdependent to allow one confidently to assign the normal distribution or its relatives to stand in for the probability distributions actually encountered".

A much more illuminating scientific evaluation can be derived from analysis of the residual measures, graphical results, and diagnostic simulations suggested here. Assessment of the adequacy of model results should be based on the summary and residual measures and graphical information, knowledge of model sensitivity to perturbations in key inputs, an understanding of the theoretical formulation and numerical implementation of the model, and knowledge of the structure and uncertainties in the various input data bases. Given the present limitations of statistical testing with sparse data sets, as applied to photochemical model predictions, reliance on confidence bands and statistical significance tests raises unnecessary risks of inappropriate rejection of adequate model results or the converse.

2.2 Graphical Procedures

Certain features of a photochemical grid model simulation are best analyzed through graphical means. In addition to revealing important qualitative relationships, graphical displays also supply quantitative information. Nine different graphical methods are suggested to display:

- The relationship between the five accuracy measures;
- The temporal correlation between predictions and observations;
- The spatial distribution of predicted concentration fields;

- The correlation between hourly pairs of predictions, observations and residuals;
- The variation in bias and error estimates as functions of time and space; and
- The degree of mismatch between model predictions and point measurements.

Each of the recommended graphical procedures is described in the following section together with examples of suggested plotting methods.

2.1 Accuracy Plot

Two accuracy plots are recommended. One depicts relationships between the five accuracy measures introduced earlier while the other plot summarizes peak prediction accuracy at all monitoring stations, revealing the presence of subregional prediction bias if it occurs.

Figure 2-2 is a histogram that displays the calculated values of A_{ts} , A_s , A_u , and \bar{A} . In the example shown, all but the unpaired accuracy estimates are negative, indicating an underestimation of peak ozone. The positive value A_u results from a model prediction in a grid cell that exceeds the highest measurement in the network by about 7 percent.

Figure 2-3a is a histogram plot showing the peak observed and predicted concentrations (unpaired in time) at each monitoring station above the cutoff concentration. Figure 2-3b contains values of A_s calculated at each monitoring station. The shaded region corresponds to the normalized gross error calculated from Equation 2-12. The dashed line in Figure 2-3b corresponds to the mean normalized bias, calculated from Equation 2-8. Both the bias and error statistics are included on the plot for easy reference. Where feasible, the ordering of the stations on the plot may correspond to the predominant wind flow direction from the primary source region to the downwind receptor locations. Thus, monitors within the high emission source region(s) are plotted on the left

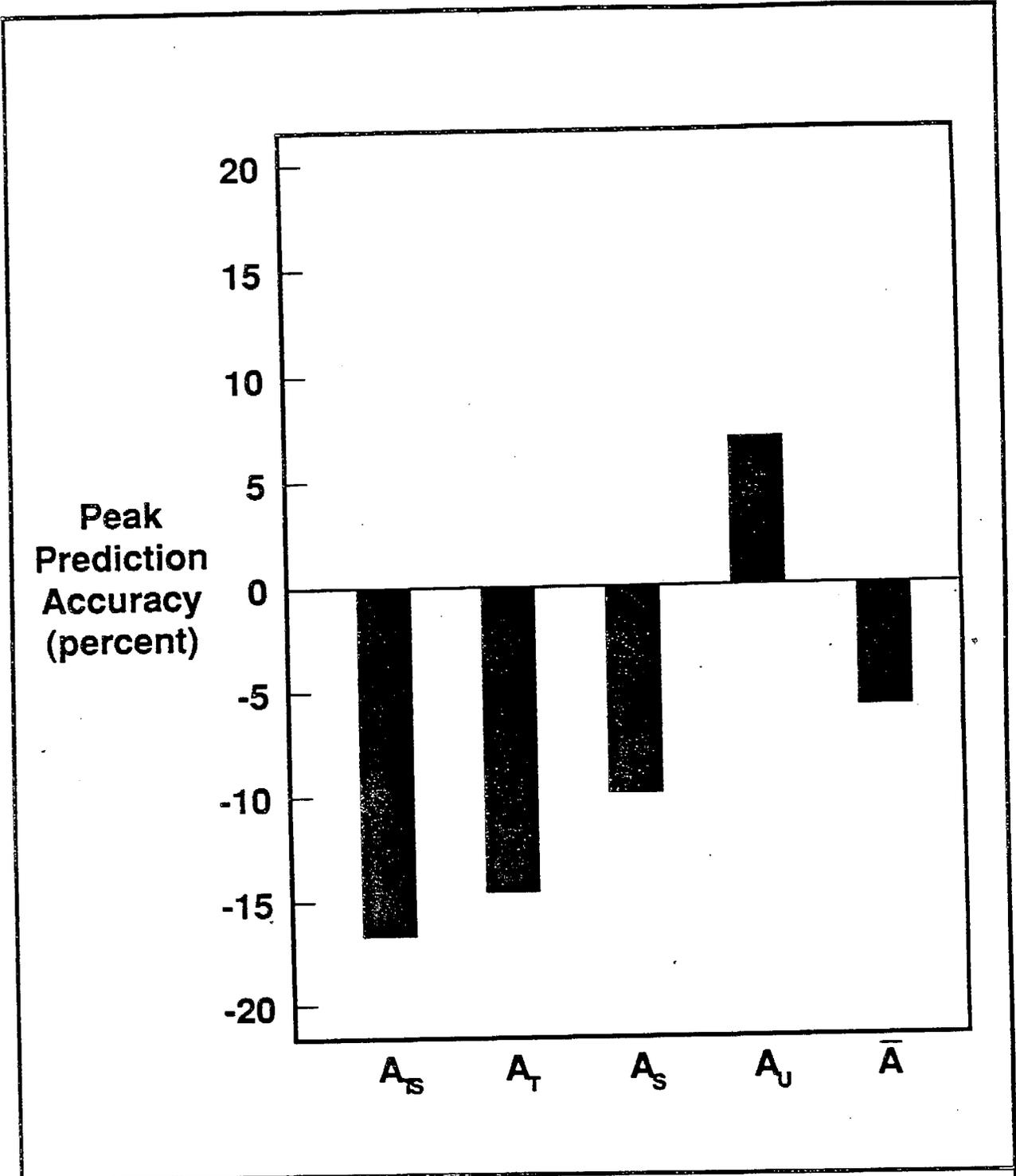
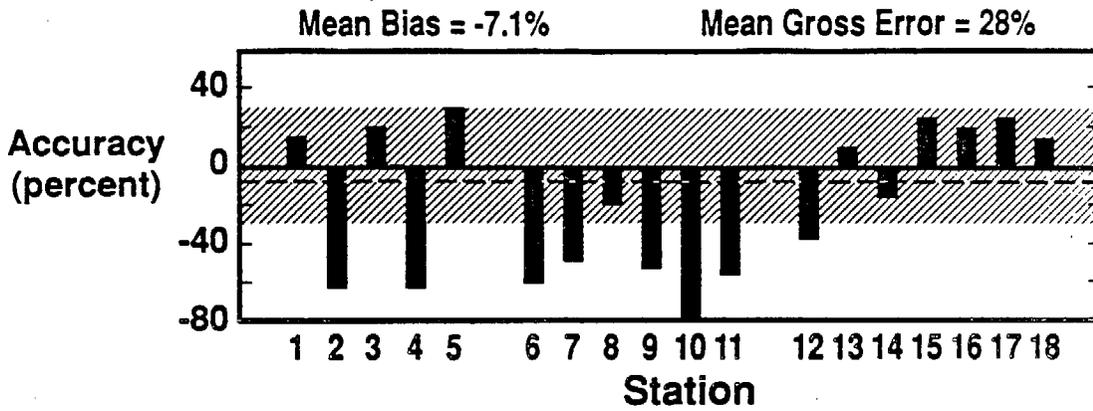
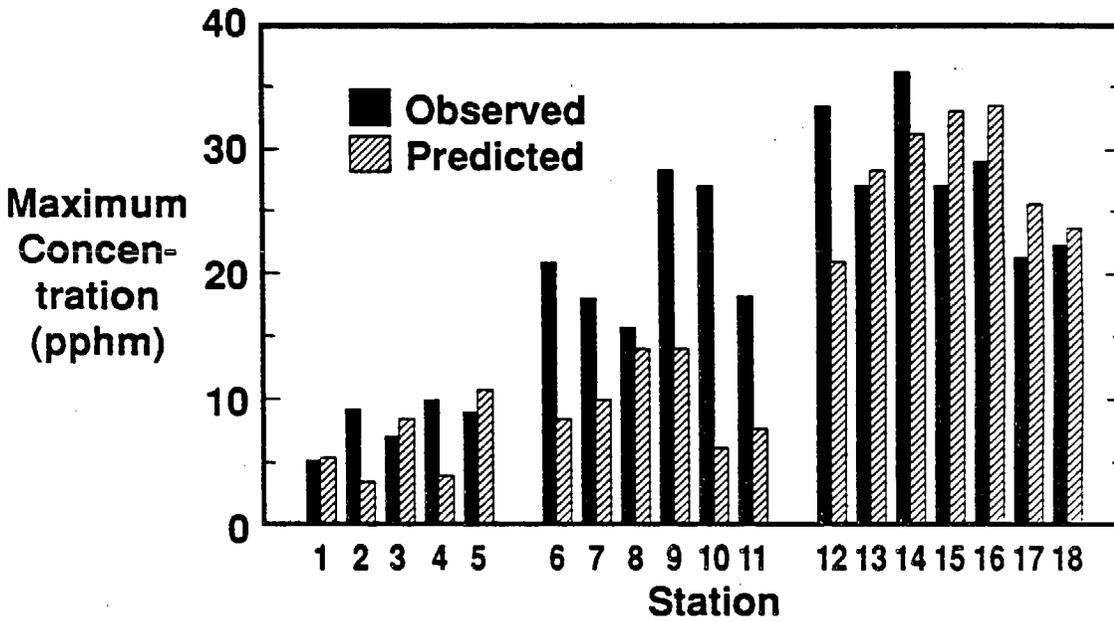


Figure 2-2. Different Measures of Peak Prediction Accuracy.



b. Accuracy of Individual Peak Predictions



a. Peak Predicted and Observed Concentrations at all Monitoring Stations

Figure 2-3. Accuracy of Station Peak Predictions.

and those in the receptor locations are plotted on the right. Alternatively, the grouping may be made by geographical subregion. Figure 2-2 reveals directly the degree of improvement in peak prediction accuracy as the requirements for temporal and spatial pairing are relaxed.

2.2.2 Time Series Plots

Probably the most useful graph for depicting photochemical model results is the time series plot. Developed for each monitoring station for which observed concentrations are available, this plot presents the hourly predictions and observations throughout the simulation period. As shown in Figure 2-4, the time series plot consists of the hourly averaged observations (the solid circles) and the hourly averaged predictions, the latter being fitted by a smooth continuous line. Here, the predictions represent the four-cell weighted average. At each hour, the absolute value of the concentration residual (Equation 2-10) is calculated and plotted as a dashed line on the same plot. Alternatively, the hourly residual may be plotted as a signed quantity, thereby providing even more information than the absolute residual, but this will require a somewhat more complicated graphical presentation. The time series plot in Figure 2-4 may be enhanced by including the average residual over the day together with the maximum predicted and observed values. These latter statistics are not shown in the Figure 2-4 example.

Time series plot are very useful. One may determine the model's ability to reproduce the peak prediction, the presence or absence of significant bias and errors within the diurnal cycle, and whether the "timing" of the predicted concentration maximum agrees with the observation. By including the residual plot on the same graph, prediction biases are more apparent.

2.2.3 Spatial Time Series Plots

Conventional time series plots do not reveal situations where the model predicts concentrations comparable in magnitude to the observations a short distance away from the monitoring station. A second time series display, called

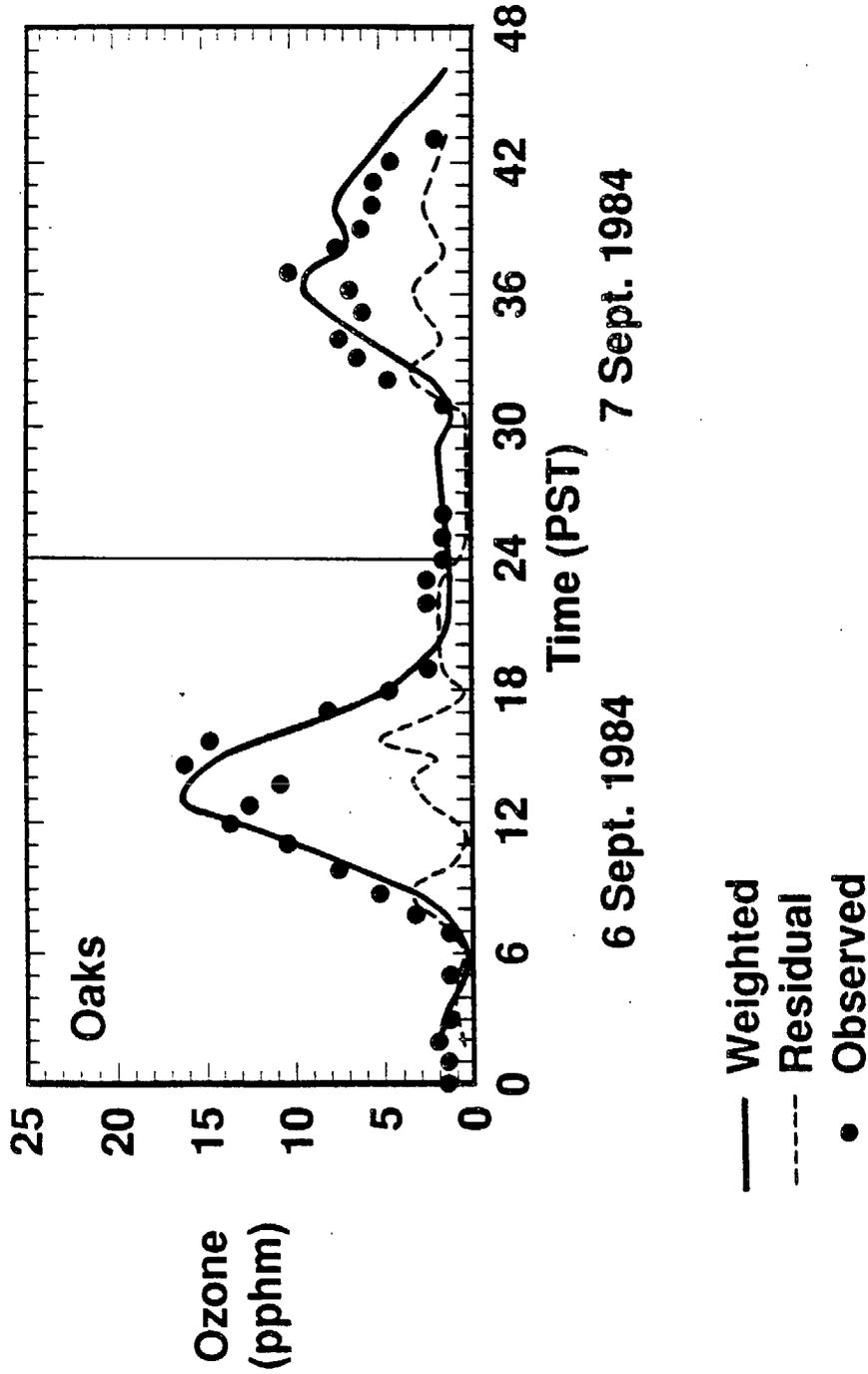


Figure 2-4. Ozone Time Series Plot at Thousand Oaks for 6 - 7 September 1984
Base Case In Ventura-Santa Barbara.

a "spatial time series plot", is recommended for this purpose. These plots provide information about the degree to which model discrepancies result from the procedure for selecting the predicted values. There is no a priori reason to select the four-cell weighted average prediction over the prediction in the specific grid cell containing the monitor (i.e., the "cell value"), or perhaps the grid cell prediction within any of the four adjacent cells that is closest in magnitude to the observed value (i.e. the "best" estimate). Thus, additional spatial time series plots may be constructed for each monitoring station by plotting the hourly observations together with three sets of model predictions:

- The four cell weighted average ("weighted");
- The prediction in the grid cell containing the monitor ("cell");
and
- The prediction closest in magnitude to the observed value, where the prediction at a given hour is drawn from one of the four nearest grid cells ("best").

Figure 2-5 gives an example of a spatial time series plot. An option with this plot (not shown) is to graph the hourly maximum difference between the largest and smallest of the three predicted values, in a manner analogous to the residual curve shown in Figure 2-4.

The spatial time series plot provide useful diagnostic information about the "steepness" of the concentration gradients in the simulated fields. Small differences between the three curves in Figure 2-5 indicate relatively flat concentration gradients. Conversely, steep gradients may produce fairly large differences between weighted, cell, and best predictions. Sometimes, calculated measures such as bias and error may not be greatly dependent upon the method of choosing the predicted value (e.g., the four-cell weighted average) and this can be determined directly from these plots. Spatial time series plots are one method of revealing the correspondence or "commensurability" between volume-averaged model predictions and point measurements. We address this subject of commensurability in greater depth in Section 4.0.

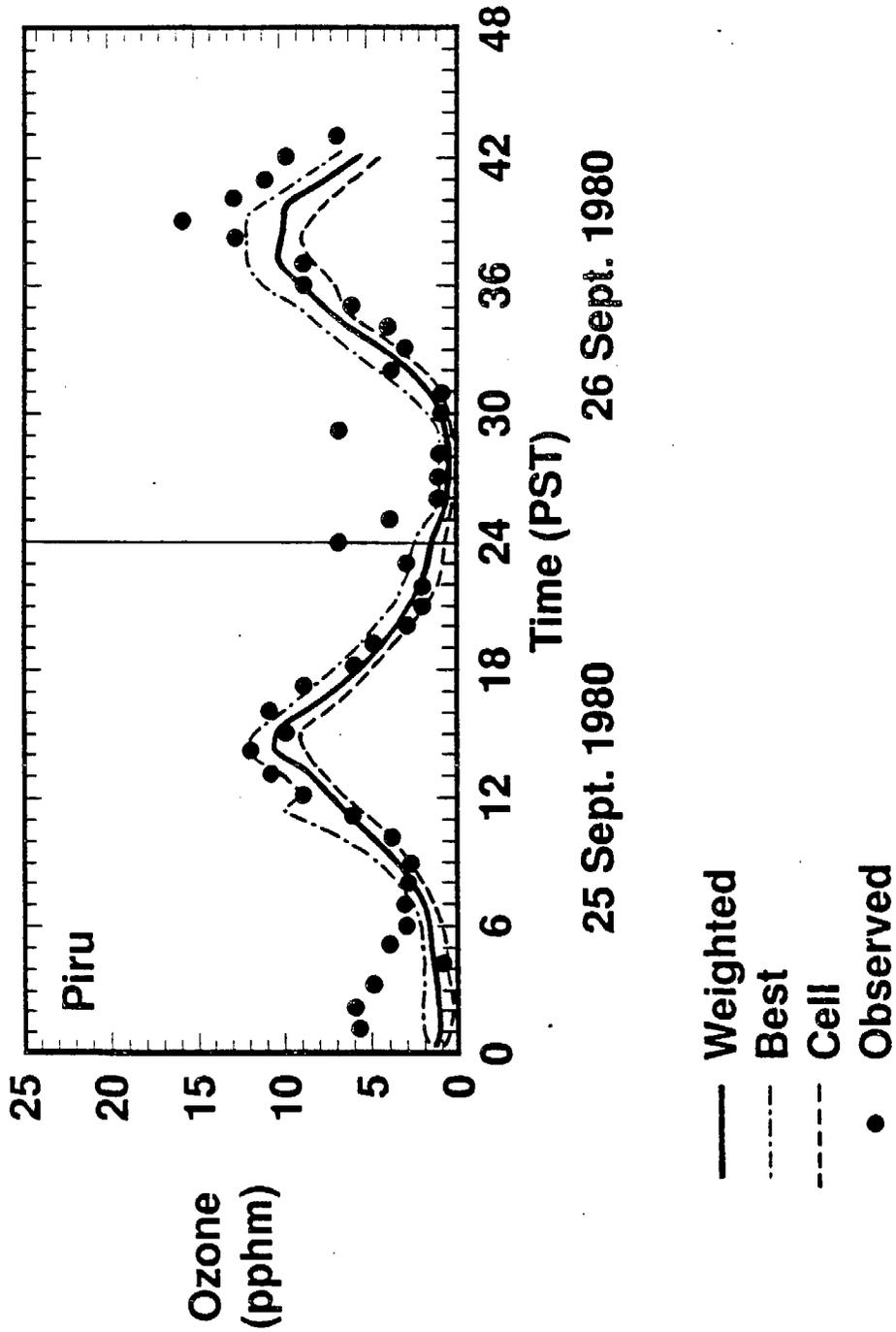


Figure 2-5. Spatial Ozone Time Series Plot at Piru for 25 - 26 September 1980
Base Case in Ventura-Santa Barbara.

2.2.4 Ground Level Isopleths

Ground-level isopleths (Figure 2-6) display the spatial distribution of predicted concentration fields for any selected hour. Combined with time series plots, these isopleths are a very useful method for investigating photochemical grid model results. The isopleth plots are developed by computer-contouring the hourly, gridded model predictions. The information content of these plots is enhanced by including the following:

- A base map identifying significant geophysical and political boundaries;
- Locations of air monitoring stations;
- The observed concentrations at each monitoring station by a bold numeral;
- The location of the peak prediction (signified by an asterisk); and
- The magnitude of the peak grid cell prediction.

These isopleths supply direct information about the magnitude and location of pollutant concentrations, and help to identify situations where sub-regional biases may be attributed to spatial misalignment of the predicted and observed concentration fields.

2.2.5 Scatter Plot of Predictions and Observations

Scatter plots are a familiar means of visually assessing the extent of bias and error in hourly prediction-observation pairs. The scatter plot in Figure 2-7 is developed by plotting all hourly-averaged prediction-observation pairs for which the observed concentration exceeds the cutoff value. Similarly, Figure 2-8 is developed from the pairs of maximum hourly predicted and observed values at each monitoring station. The predicted maximum is the highest value

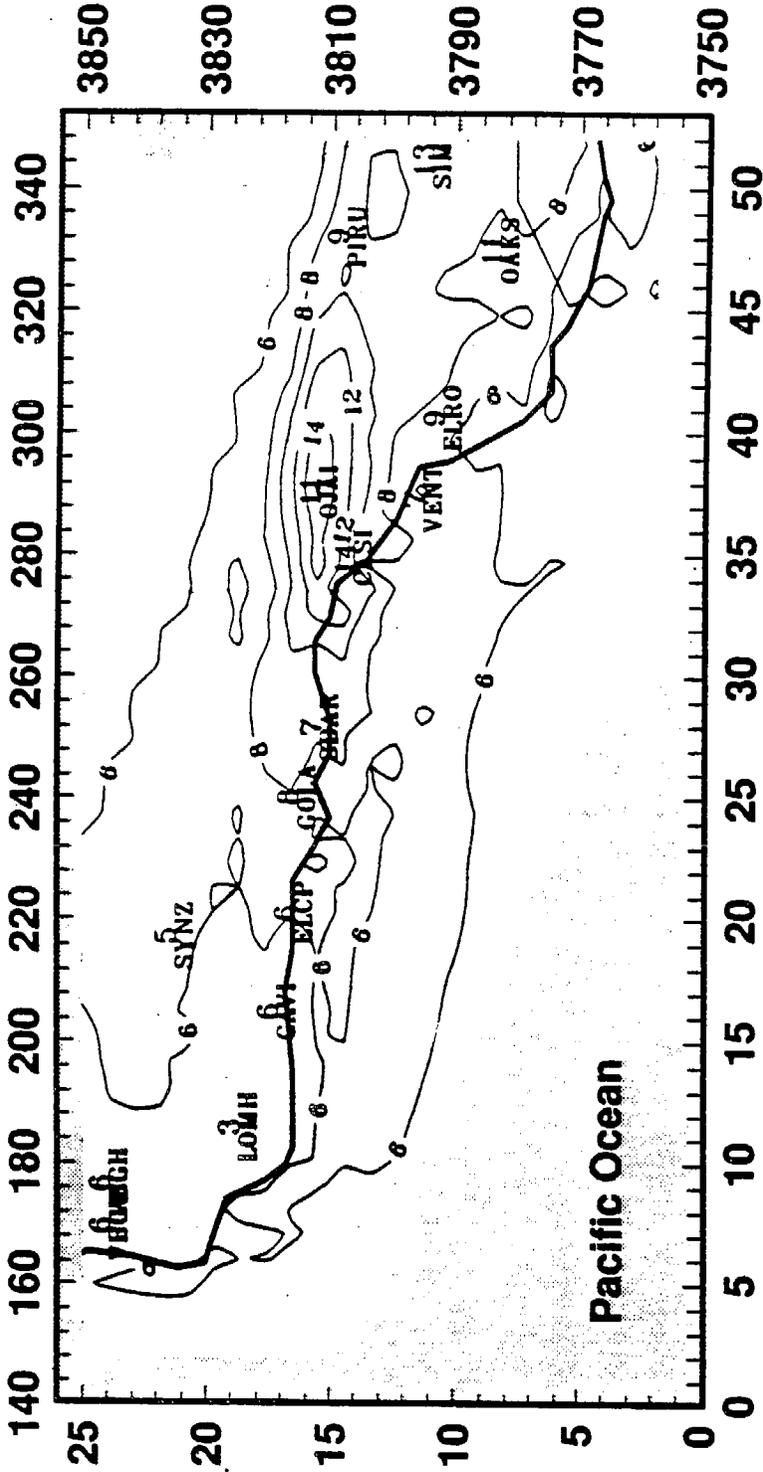


Figure 2-6. Ground Level Hourly Average Ozone Concentration Isopleth on 17 September 1984
Between the Hours of 1500 and 1600 PST (ppb).

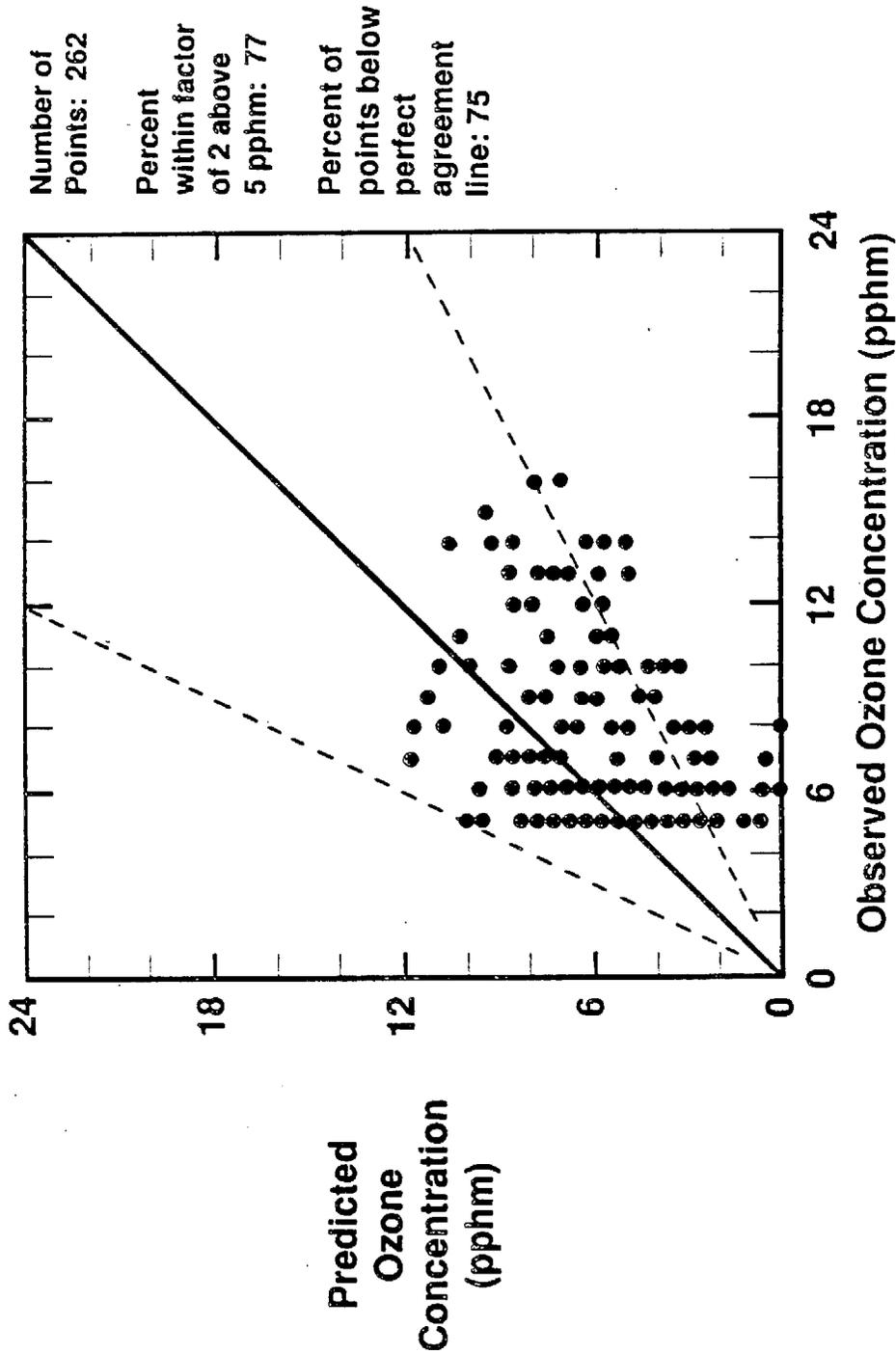


Figure 2-7. Correlation of Hourly Average Predictions and Observations at all Monitoring Stations.

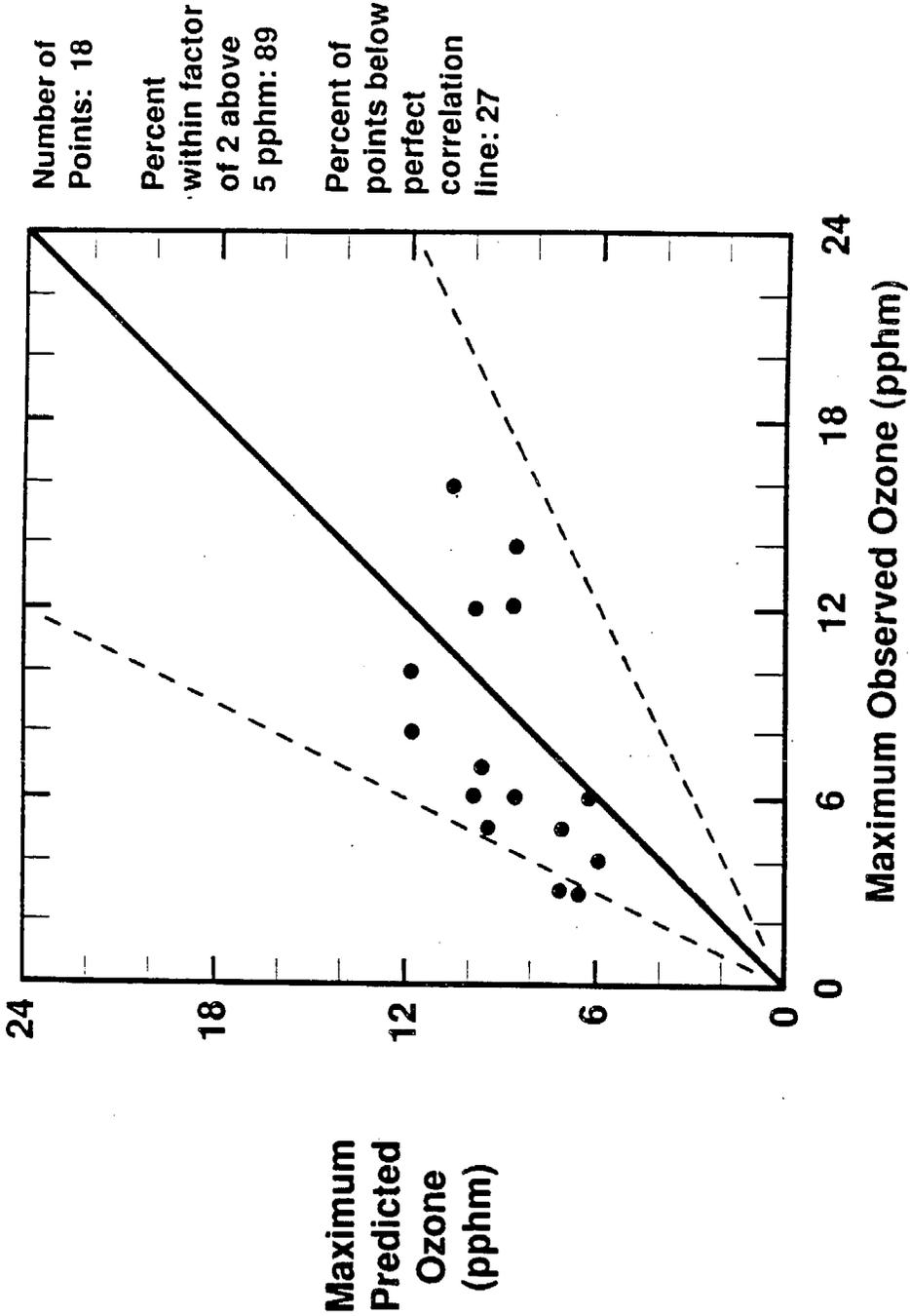


Figure 2-8. Correlation of Daily Maximum Predictions and Observations at all Monitoring Stations.

simulated within three hours of the observed maximum. In these plots, the solid diagonal line is the perfect correlation line and the dashed lines enclose the region wherein predictions and observations agree to within a factor of two.

The scatter plot gives a quick visual indication of the extent of over- or underestimation in the hourly predictions and whether there appear to be noticeable nonlinear relationships between predictions and observations over the concentration range studied. Bias is indicated by the preponderance of data points falling above or below the perfect correlation line. The dispersion (spread) of points provides a visual indication of the general error pattern in the simulation. Scatter plots help identify outlier prediction-observation pairs, i.e., a seemingly discrepant prediction-observation pair that may result from erroneous data, a fundamental flaw in the model, or some other cause that requires investigation. These plots provide little diagnostic information about sub-regional performance problems, temporal or spatial misalignments, or other inadequacies in the simulation. In addition, scatterplots do not reveal the temporal correlation between various prediction-observation pairs.

2.2.6 Scatter Plot of Residuals and Observations

Residual scatter plots reveal the distribution of hourly average model discrepancies (positive and negative) as a function of concentration level. This is a graphical display of the data elements that make up the bias and error calculations. Hourly concentration residuals for all monitoring stations are plotted as a function of observed concentration for all pairs above the cutoff value. Figure 2-9 depicts the hourly residual scatter plot. The daily maximum residual plot (Figure 2-10) is based on data pairs involving the maximum observed concentration at a monitor station and the maximum predicted value at the same station within three hours of the peak.

Residual scatter plots describe prediction discrepancy throughout the observed concentration range. The plot does not reveal the existence or causes

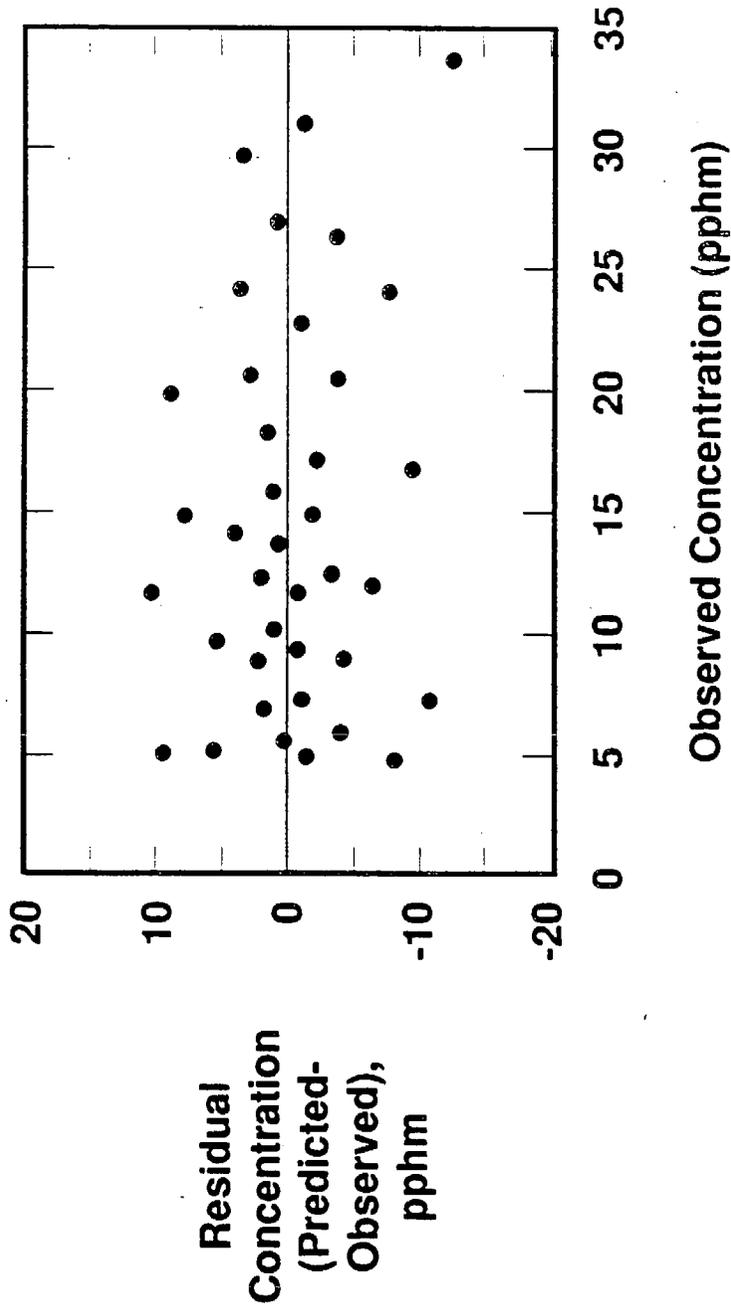


Figure 2-9. Distribution of Hourly Concentration Residuals as a Function of Observed Concentration.

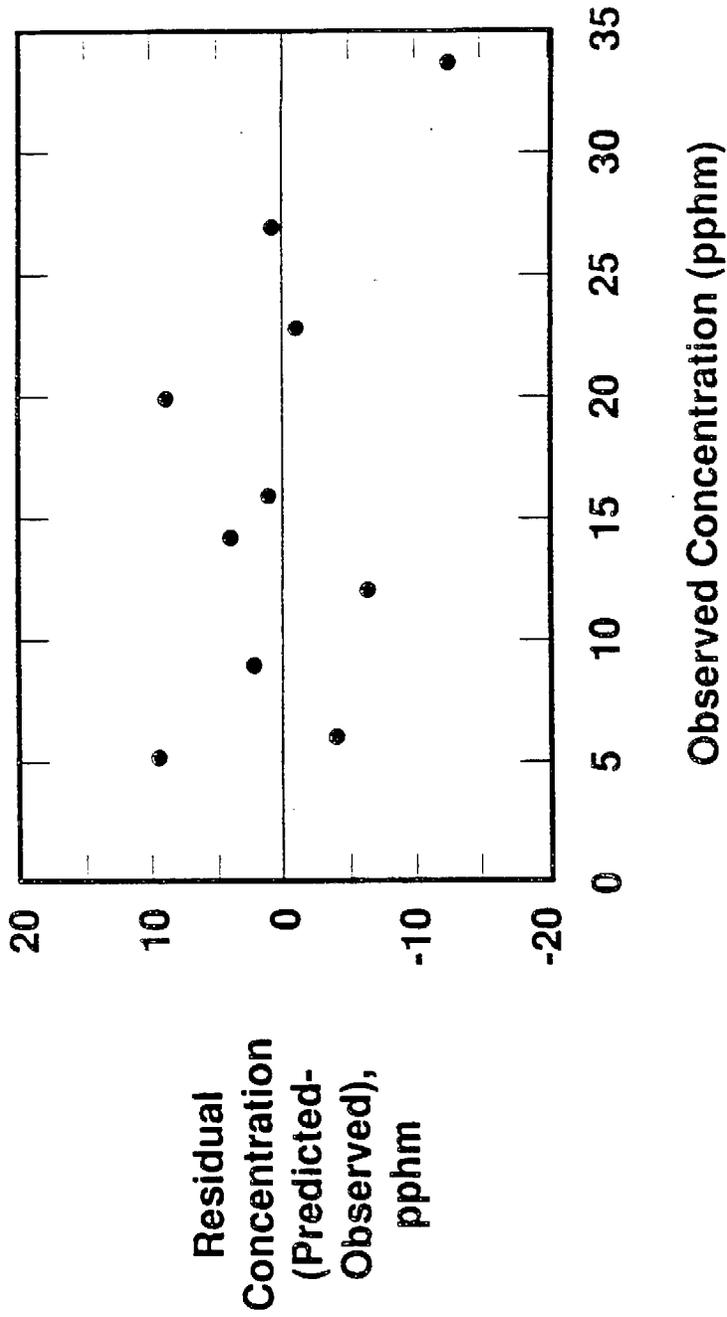


Figure 2-10. Distribution of Maximum Concentration Residuals as a Function of Observed Concentration.

of sub-regional or timing performance problems. Absence of bias is suggested by no systematic tendency for the data points to fall above or below the ordinate; however, as noted previously, important subregional biases may still exist in the presence of a zero overall bias estimate.

2.2.7 Bias Stratified by Concentration

The bias-concentration plot, derived from the residual scatter plot in Figure 2-9, depicts the degree of systematic bias in hourly-averaged model predictions (paired in time and space) as a function of observed concentration level. This plot (and the companion error-concentration plot) have proven to be among the most useful graphical aids in diagnostic model evaluation. Figure 2-11 shows a bias-concentration plot. The observed concentration range is divided into several equal-sized concentration bins and the normalized bias within each bin is calculated from Equation 2-8. The bin average biases are then plotted as a function of concentration level. Although not shown in Figure 2-9, one may include the mean normalized bias statistic on the plot for easy reference.

The bias-concentration plot reveals the existence of under- or over-estimation throughout the concentration range. The frequent situation of under-estimations at the low concentration end and overprediction at the high end, masked in the overall bias statistic (Equation 2-8), are revealed directly in Figure 2-11.

2.2.8 Gross Error Stratified by Concentration

The gross error-concentration plot, also derived from the residual distribution, depicts the degree of error in model prediction (paired in time and space) as a function of observed concentration level. The observed concentration range is divided into several equal-sized concentration bins. Then, average value of the normalized gross error (Equation 2-12) within each bin is calculated. Bin averages are plotted as a function of the observed concentration level as shown in Figure 2-12. One may wish to print the mean normalized gross error in Figure 2-12 for easy reference.

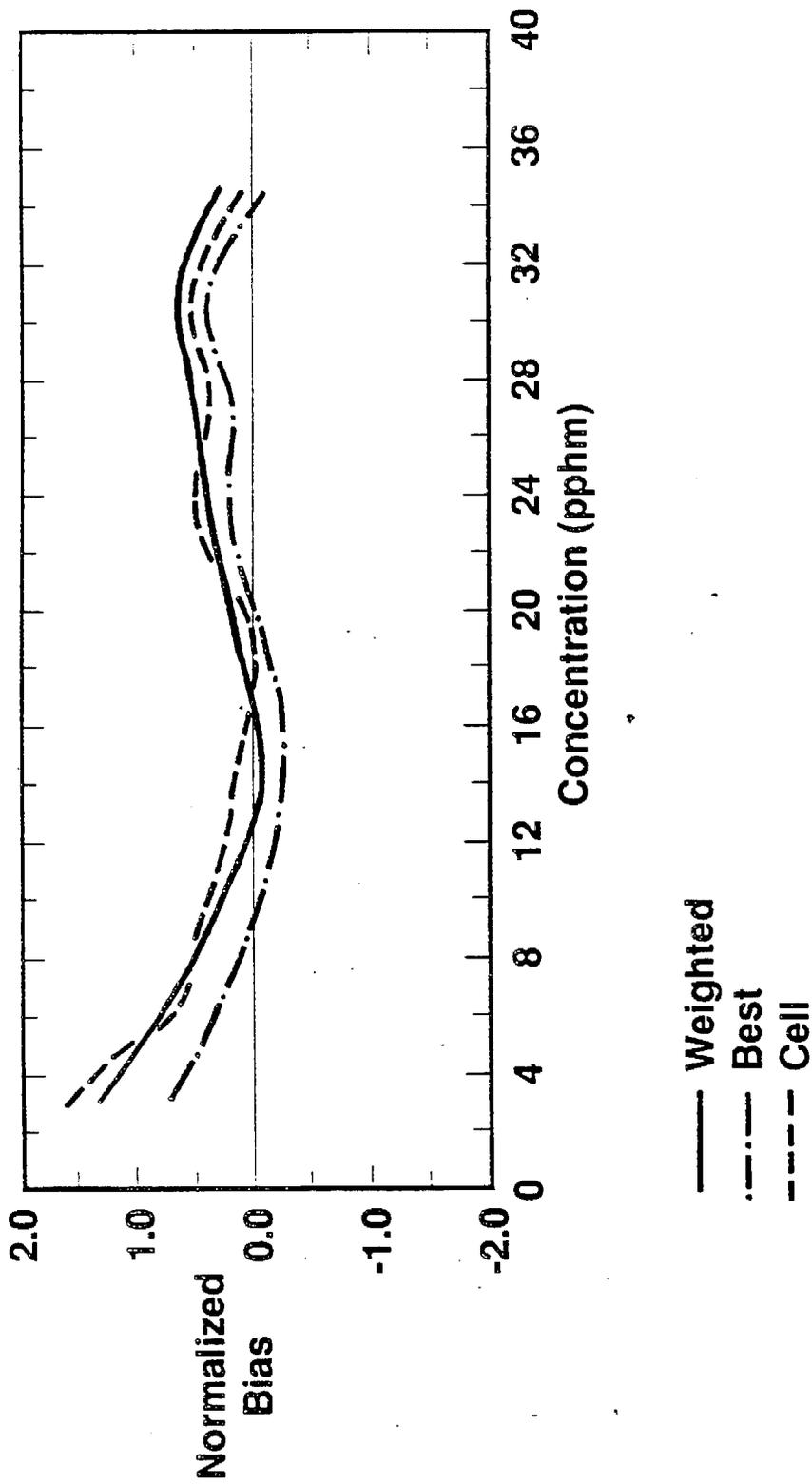


Figure 2-11. Bias as a Function of Observed Concentration Level.

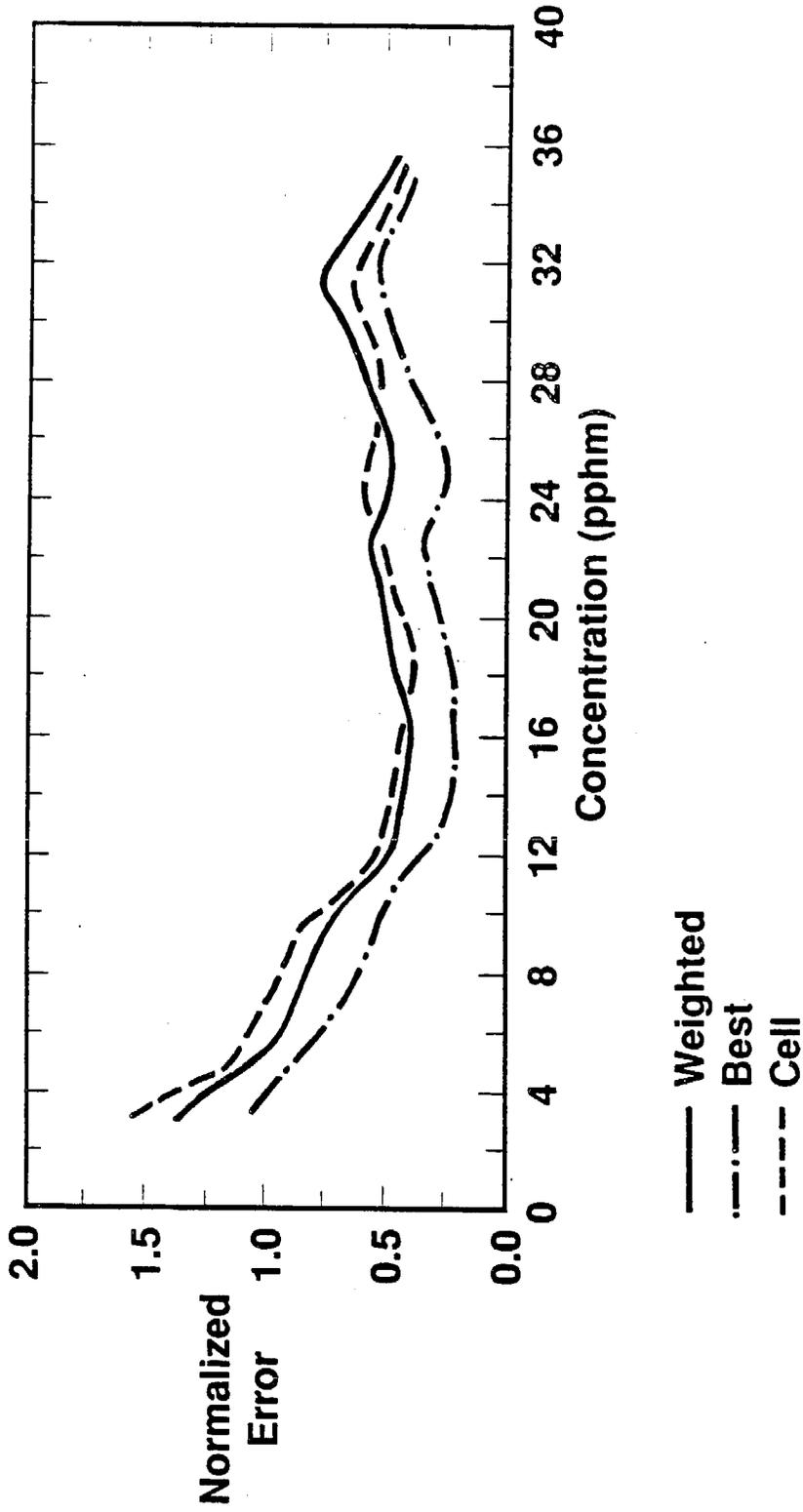


Figure 2-12. Gross Error as a Function of Observed Concentration Level.

The gross error-concentration plot reveals the variation in model error at various intervals throughout the concentration range. The plot must be interpreted carefully, however, remembering that the concentration residual is normalized by the observed value.

2.2.9 Bias Stratified by Time

The bias-time plot identifies specific time periods within the photochemical simulation when systematic tendencies toward under- or overestimation occurs. The bias-time plot is constructed in a manner similar to the bias-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. An example of this plot is shown in Figure 2-13.

Systematic bias in model predictions during specific periods within the diurnal cycle may have several causes: microscale chemical interactions between fresh emissions and ambient concentrations; biases in vertical mixing or wind transport; "timing" problems with the chemistry; non-representative temporal distributions assumed in the emissions inventory, and so on. While the bias-time plots may not clearly pinpoint the causes of bias, they do aid in defining the time intervals when the bias is most apparent. This helps focus subsequent diagnostic investigations. Also, use of the three methods for reporting the model prediction (e.g. weighted, best, and cell) may provide some information about the spatial extent of the apparent bias.

2.2.10 Gross Error Stratified by Time

The gross error-time plot (Figure 2-14) identifies specific time periods when gross errors in the model predictions may be a problem. This plot is constructed in a similar manner as the error-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration. When interpreting the gross error-time and bias-time plots, the analyst must remember that the concentration levels of all pollutants vary throughout the diurnal cycle. For example, nighttime and midday normalized

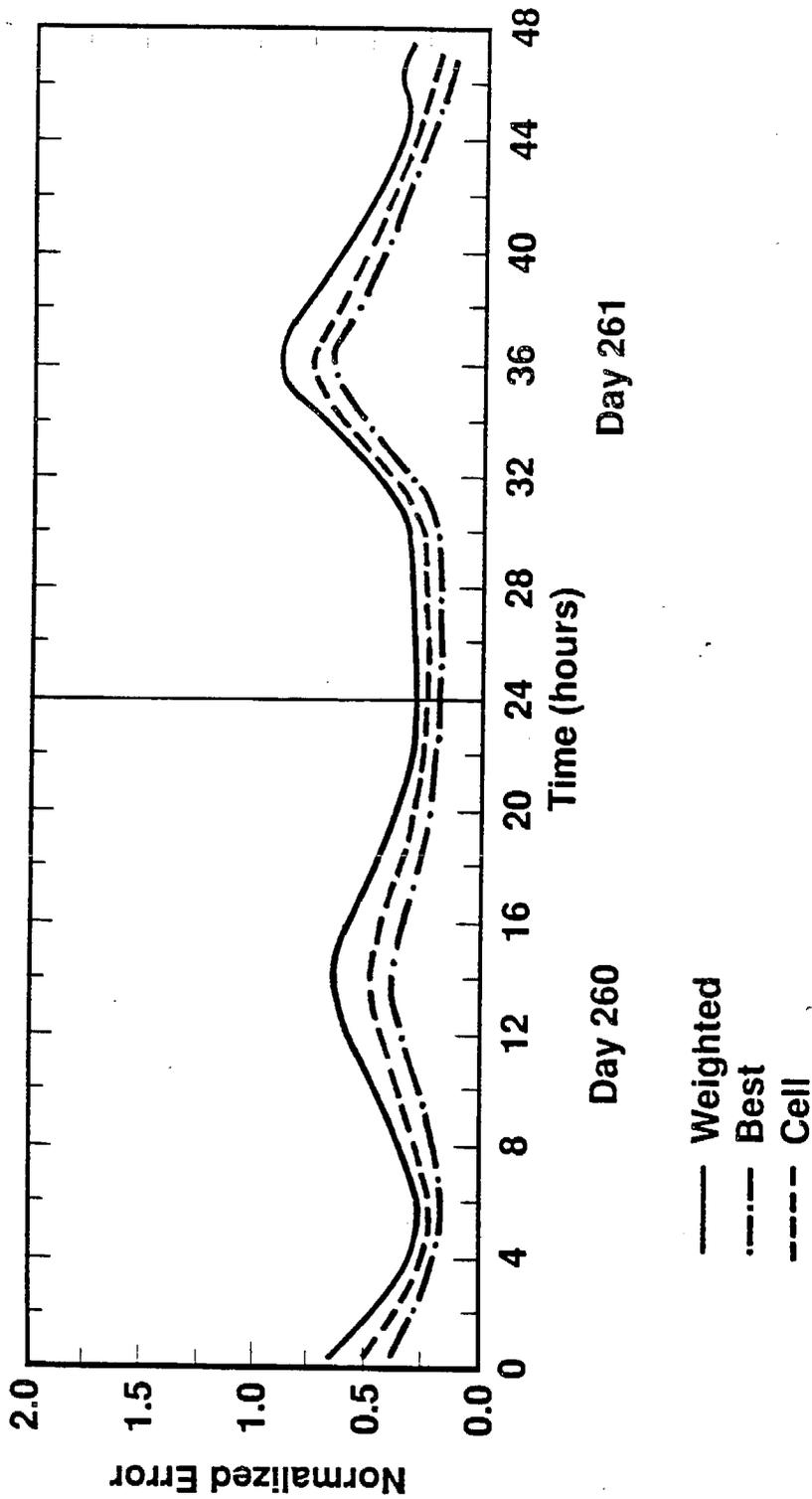


Figure 2-13. Error as a Function of Simulation Time.

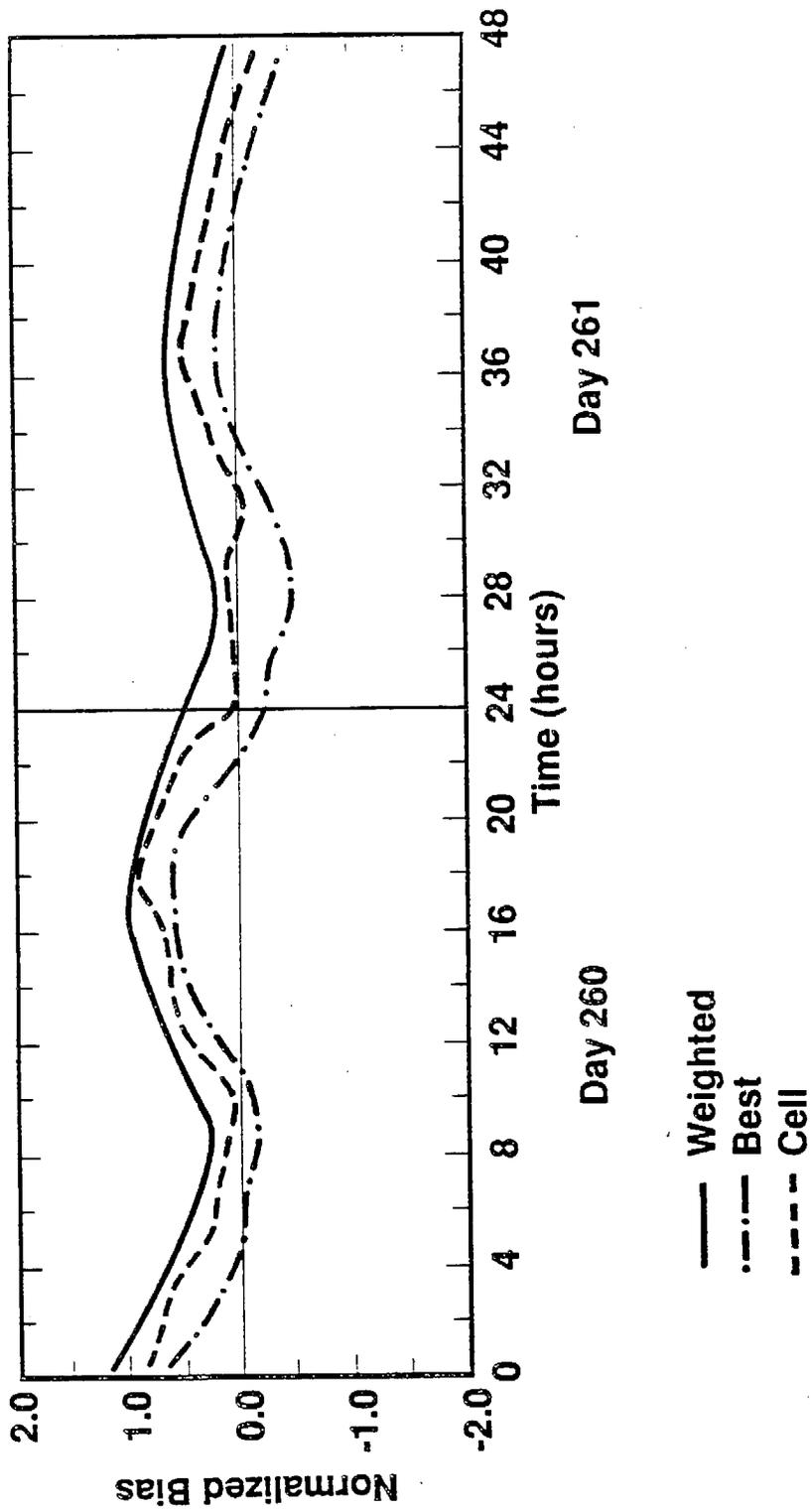


Figure 2-14. Bias as a Function of Simulation Time.

gross errors of comparable magnitude will generally represent different concentration residuals.

To aid diagnostic analysis, one may develop bias and error plots similar to Figures 2-11 through 2-14 in which several different species are plotted together. For example, the gross error in NO, NO₂, and ozone concentration may be plotted as a function of concentration level or simulation time on the same plot. This technique facilitates evaluation of model performance for multiple species, as discussed in greater detail in Section 3.0.

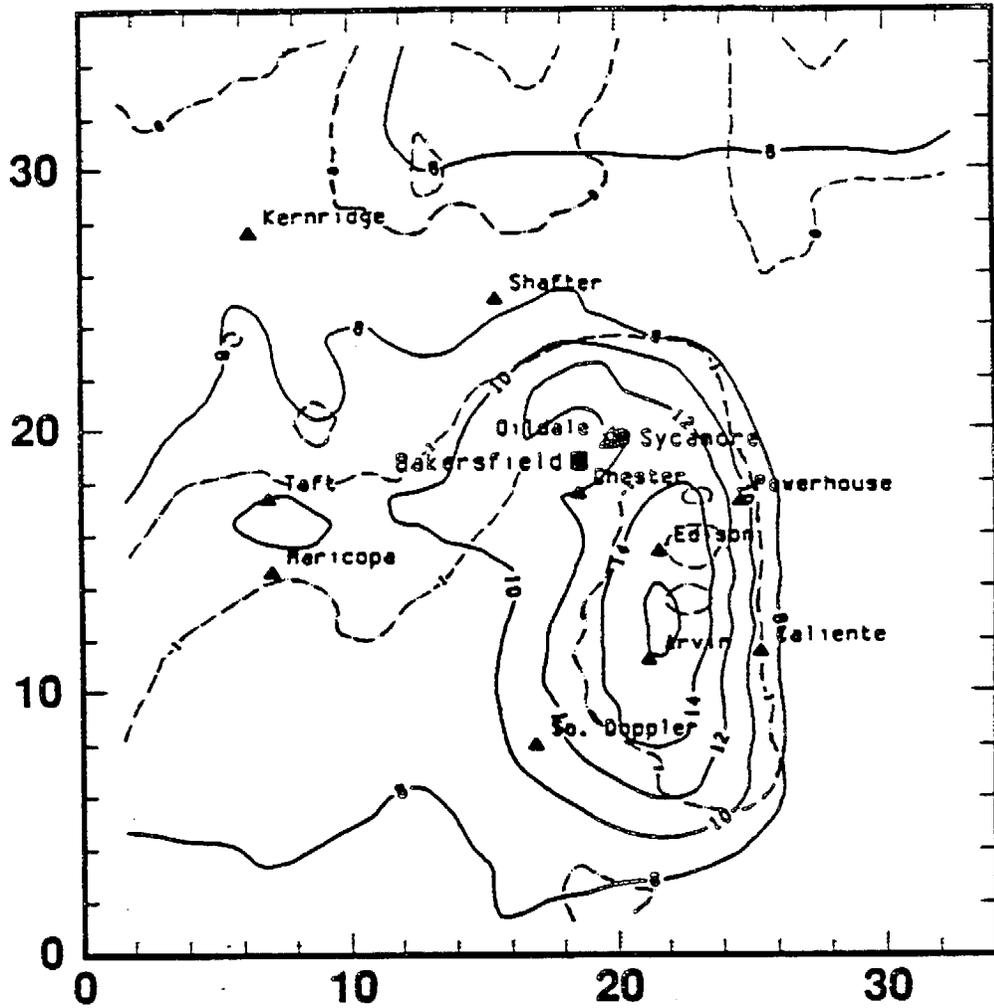
2.3 Diagnostic Simulations

Six diagnostic simulations are suggested as necessary to accompany the numerical and graphical comparisons of predictions and observations previously outlined. These are outlined below.

2.3.1 Zero Emissions

The main purpose of the zero emission simulation is to ensure that the base case simulation results are influenced appropriately by the emissions inputs. Zeroing out all emissions should lead to much lower reactive species concentrations on the second and subsequent simulation days. The zero emissions simulation is performed by exercising the base case run with all emission values reduced to zero. All other model input files remain unchanged from the base case.

The results of the zero emissions simulation, and the other diagnostic simulations, may be presented in three ways. First, a so-called Deficit-Enhancement (D-E) plot is developed for each simulation hour of interest. The D-E plot is developed by subtracting the hourly-average gridded diagnostic run concentrations from those of the base case and plotting the residuals as isopleths. The dashed lines in Figure 2-15 demonstrate this technique which has been applied to a simulation involving only moderate emissions reductions. For handy reference, the base case concentrations for the same hour may be plotted as solid contour lines. Figure 2-16 gives the hourly time series plots for the diagnostic run.



--- Concentration decreases (pphm)
— Base case predictions

Figure 2-15. Deficit-Enhancement Plot of the Effects of Moderate Emissions Reductions on Ground Level Ozone Concentrations - 1600 PST on 6 September 1984.

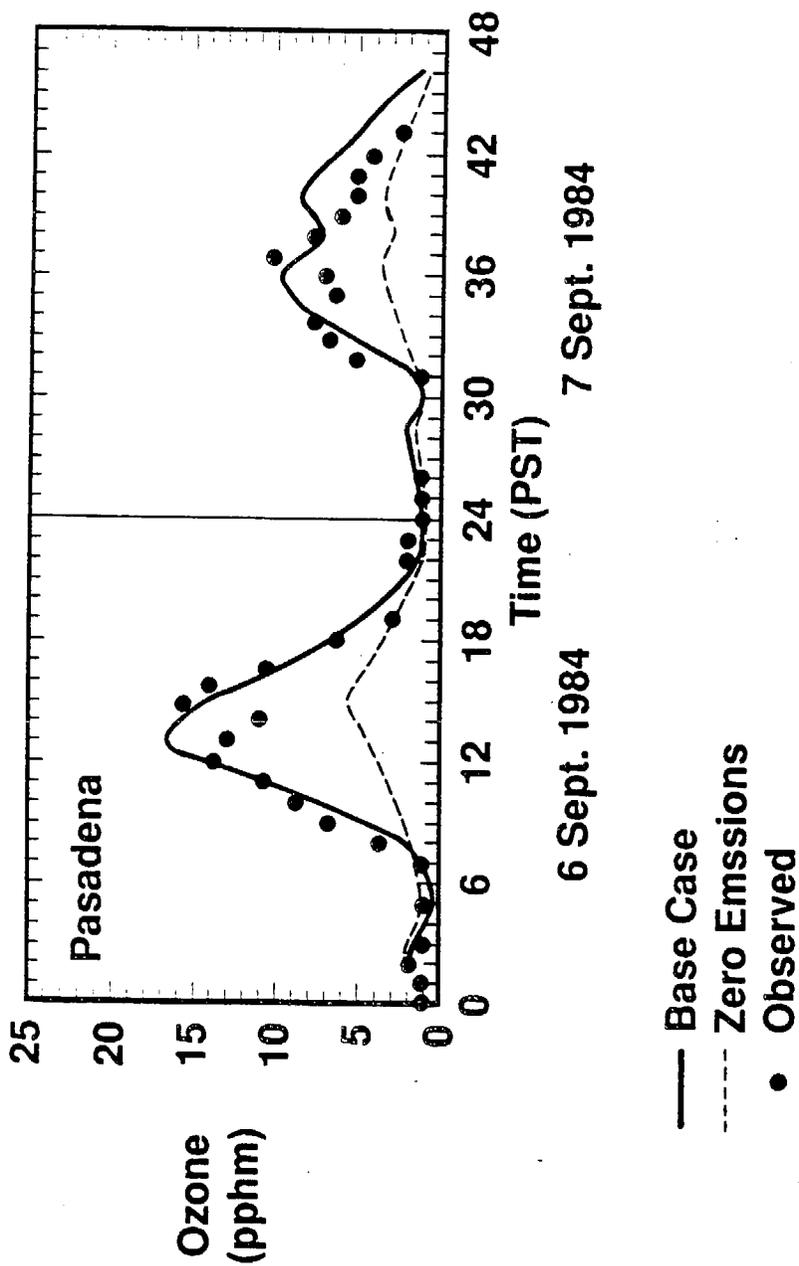


Figure 2-16. Time Series Plot of Base Case and Zero Emissions Diagnostic Run at the Pasadena Monitoring Station.

It is constructed with the base case predictions represented by a solid line and a zero emissions diagnostic run results depicted by a dashed line. Finally, Table 2-1 lists all monitoring stations and the maximum predicted concentrations (unpaired in time) in the base case and a moderate emissions reduction diagnostic run.

The zero emissions diagnostic run should produce significantly-reduced concentrations, close to background or to the levels reflective of the inflow boundary conditions. If not, there is reason to question the adequacy of the simulation. Lack of sensitivity to emissions may indicate inappropriately high initial conditions, improper boundary conditions, or some flaw in the model itself. Quite apart from these concerns, insensitivity to emissions raises serious questions about the usefulness of the simulated episode for control strategy development and assessment.

2.3.2 Zero Initial Conditions

The zero initial conditions simulation reveals how much of the second (or third) day predictions result from the initial field used to start the simulation. This simulation is performed by setting all initial concentrations to zero. This procedure may cause numerical instabilities; where this occurs, the initial field should be reduced to values as close to zero as possible, certainly below those of global background or inflow boundary conditions.

Deficit-enhancement and time series plots may be used to display the results of this simulation. If the initial field is completely "washed out" of the model domain by the second or third day, the D-E plots will indicate essentially no differences between the diagnostic and base case runs on the following day(s). For stagnation episodes, some residual effects of initial conditions may be seen even on the third day of a multiple-day simulation (Tesche and McNally, 1989).

TABLE 2-1. MAXIMUM CONCENTRATIONS FOR BASE CASE AND MODERATE EMISSIONS REDUCTION DIAGNOSTIC SIMULATIONS

(concentrations in pphm)

Monitoring Stations	Base Case Maximum Concentration	Diagnostic Simulation Maximum Concentration	Ratio of Diagnostic to Base Case Concentrations	Percentage Difference
El Capitan B	6.0	6.3	1.05	5
El Rio	8.0	6.3	0.79	-20
Ventura	9.0	6.6	0.74	-27
Goleta	7.0	6.0	0.86	-14
Grace	8.0	7.5	0.94	-6
Hondo	6.0	7.4	1.23	23
Jalama	9.0	6.4	0.71	-29
Casitas	9.0	7.6	0.84	-16
Lompoc	5.0	6.3	1.27	27
Moorpark	4.0	7.9	1.98	98
Ojai	12.0	7.3	0.61	-39
Piru	16.0	8.8	0.55	-44
Pt. Conception	9.0	6.9	0.77	-23
Hueneme	9.0	6.8	0.76	-24
Santa Ynez	9.0	5.8	0.64	-36
Santa Barbara	9.0	6.2	0.69	-31
Simi	12.0	7.1	0.60	-40
S. Mountain	9.0	7.9	0.88	-12
State Street	4.0	5.9	1.48	48

2.3.3 Zero Boundary Conditions

The zero boundary condition simulation examines the influence of boundary values on second (or third) day concentrations, particularly in regions where the base case predictions are highest. This simulation helps identify situations where the base case results are "driven" by the boundary conditions. The zero boundary conditions simulation is performed by setting all inflow and outflow boundary values to zero. This includes the top surface of the modeling region. If this procedure leads to numerical instabilities, the boundary conditions should be set as low as possible, certainly lower than background levels. Also, the photostationary state relation (Seinfeld, 1986) that approximates the NO, NO₂, and ozone reaction cycle may be used to define boundary concentrations.

Deficit-enhancement and time series plots may be used to display the simulation results. Unless the episode is known to involve interbasin transport, the D-E and time series plots should reveal little impact of the boundary conditions in the interior of the computational domain.

2.3.4 Zero Surface Deposition

The zero deposition simulation addresses the influence of dry surface deposition removal on primary and secondary species concentrations. The zero deposition diagnostic run is exercised by setting deposition velocities for all species to zero and re-running the base case simulation. Depending on the photochemical model used, this may entail either a simple change to a model input file or an actual modification to the source code.

Deposition tests have not been reported in previous model evaluation studies so the analyst has little historical information at present to serve as a guide in interpreting the results of this investigation. Nevertheless, some general guidelines can be suggested. For primary species such as NO_x and VOCs, the downwind concentration fields should increase relative to the base case in a manner consistent with the deposition velocities for each primary species.

For reactive species such as ozone, increases and decreases may occur in the D-E contour fields depending upon the interaction between ozone and NO fields.

2.3.5 Mixing Height Variations

Mixing heights have a direct and often significant influence on ozone concentrations. The objective of the mixing height diagnostic simulation is to reveal the degree to which ozone concentrations are influenced by the height of the mixed layer. At a minimum, one diagnostic run is suggested in which the hourly mixing height values are uniformly increased by 50 percent above the base case values. This percentage increase is somewhat larger than the expected uncertainty in estimates of mixing heights typically encountered. Therefore, this simulation should provide a bound on the change in ozone predictions resulting from uncertainties in this input. Increased mixing heights typically reduce ozone concentrations although the reduction is less than a one-to-one change. One might choose, instead, to reduce the hourly mixing heights by 50 percent. The resultant increase in ozone concentrations under this scenario will typically be comparable in magnitude but of opposite sign as those for the mixing height increase case.

The results of this diagnostic simulation can be presented graphically with D-E plots (Figure 2-15) or with conventional time series plots (Figure 2-16). A more informative plot, referred to as a "time-space uncertainty plot" can be developed, although it requires carrying out several mixing height simulations instead of just one. An example of a time-space uncertainty plot is given in Figure 2-17; it was developed as follows. The solid bold line in the figure represents the base case ozone predictions at a particular air monitoring station. The boxes indicate the observed ozone concentrations at each hour and the vertical lines associated with each box represent the estimated uncertainty of the ozone measurement. The magnitude of these measurement uncertainties has been exaggerated in Figure 2-17 in order to exemplify the method. These estimates could include a component related to the spatial representativeness of the monitoring station if one is able to develop it. The

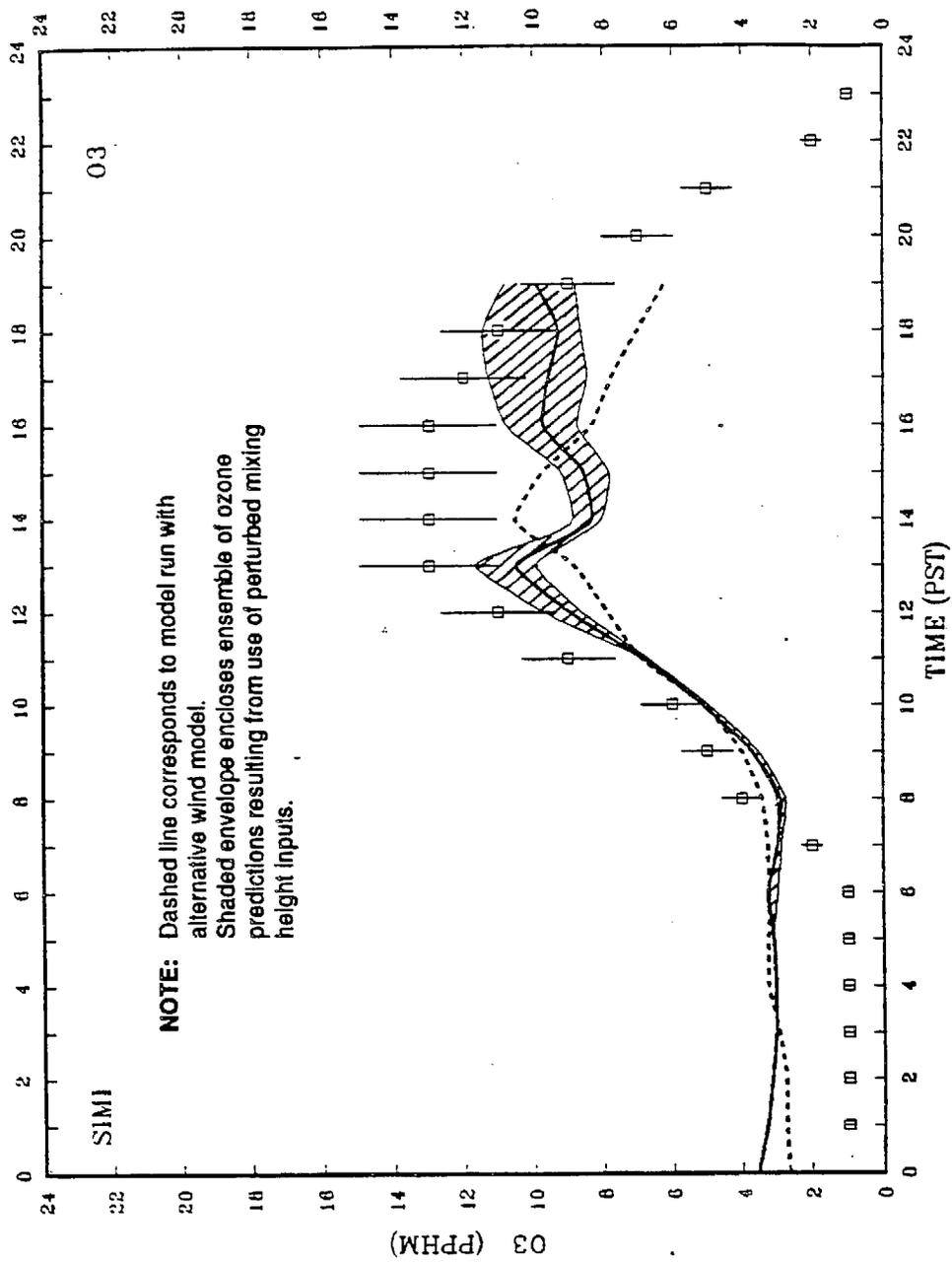


Figure 2-17. Time-Space Uncertainty Envelopes of Ozone Prediction and Measurement Uncertainties on 17 September 1984 at the Simi Monitoring Station.

solid lines enclose an ensemble of time series profiles obtained from several diagnostic runs involving different increases and decreases in the base case mixing heights. In the example shown, these mixing height uncertainties were derived from more than a dozen simulations of a numerical mixing height model (Tesche, McNally, and Wilkinson, 1988). Ideally, the ensemble of photochemical model predictions (enclosed by the thin solid lines) would trace a path within the upper and lower uncertainty bounds of the hourly ozone measurements. Comparisons between the estimated measurement uncertainties and the variation in ozone predictions due to mixing height variations could be used to guide subsequent activities aimed at additional data collection, model refinement or both. Finally, the dashed line in Figure 2-17 represents an additional ozone simulation in which an alternative procedure for constructing wind fields was used. Thus, the time-space uncertainty plot may be used to compare directly the results of different diagnostic runs.

2.3.6 Reduced Wind Speeds

The final diagnostic simulation we recommend entails a 50 percent reduction in the magnitude of the winds input to the photochemical model. More complex and potentially very insightful diagnostic runs focusing on windfields might be conceived (e.g. random perturbation of grid-point speeds and directions) but an initial characterization of the ozone model's sensitivity to ventilation is an appropriate starting point. This simulation is performed by decreasing all grid-point wind components by 50 percent and re-running the photochemical model. Based on previous windfield diagnostic results, the magnitude of the hourly ozone concentrations, including the peak value, should increase relative to the base case although the percentage increase should be less than proportional to the wind speed reduction. Results of the windfield diagnostic simulation can be displayed through D-E plots or by various time series plots.

2.4 Guidelines for Assessing Simulation Results

Decision-makers and regulatory agencies seek quantitative performance standards by which any new model evaluation study may be judged either acceptable or rejected for not meeting pre-established requirements. However, we do not

endorse setting rigid criteria for model "acceptance" or "rejection" for both technical and policy reasons. Fox (1981) and Roth et al., (1988) discuss the problems with setting model performance standards for air quality models. Even after a decade of modeling research and applications studies, no scientifically sound proposals have been advanced for setting performance standards that clearly and consistently discriminate between acceptable and unacceptable model simulations. As a result, we suggest an approach that gives regulatory agencies guidance in making acceptance or rejection decisions on a case-by-case basis.

Each photochemical modeling episode exhibits distinctive aerometric and emissions features. The available data base also is unique in the amount and quality of observations available to support model evaluation and testing. In addition, the particular set of modeling procedures and codes make each application distinctive. Therefore, automatic use of standardized acceptance or rejection standards raises the risk of accepting a model evaluation that gives seemingly "good" performance statistics but for the wrong or misleading reasons or of rejecting a model evaluation that violates preset criteria for reasons related to input inaccuracies rather than fundamental flaws.

Instead of prescribing fixed performance standards, we suggest the following approach. From over 15 years of photochemical model development and testing, photochemical grid model simulations generally produce peak (unpaired) prediction accuracy, overall bias, and gross error statistics in the approximate ranges of 15-20 percent, 5-15 percent, and 30-35 percent, respectively. A study following an approved Ozone Modeling Protocol that falls below all of these ranges would not be rejected unless evidence from the model diagnostic simulations and the other numerical measures and diagnostic tests suggest unusual or aberrant behavior. For model simulations falling within these ranges, some additional diagnostic analyses (such as those proposed in Sections 3.0 and 4.0) may be appropriate to lend further support to the contention that the simulation is acceptable. For model results outside any one of these general ranges, it should be incumbent on the modeler to explain why the performance is poorer than that commonly achieved in similar applications. The modeler should also explain whether the causes of poorer performance will adversely affect the use of the model in control strategy evaluations. This methodology provides reviewing

agencies with a general model performance target, but still guards against the inappropriate rejection of less accurate model simulations when appropriate and explainable reasons can be provided.

3.0 DIAGNOSTIC MODEL PERFORMANCE EVALUATION PROCEDURES

There are at least three main reasons for carrying out the diagnostic model evaluation procedures developed in this section. These are to: (1) determine the causes of failure of a flawed model, (2) stress a model to ensure failure if indeed the model is flawed, and (3) provide additional insight into model performance beyond that supplied through the operational evaluation procedures previously introduced. Frequently, the operational model evaluation measures may not convey sufficient information about the model and data base to allow their use in emission control strategy development and testing. Particularly in complex air basins such as the South Coast and the Bay Area, the model evaluation process should be supplemented with additional diagnostic analyses that probe further into the comparisons between prediction and observation, attempting to ensure that the ozone response given by the model is correct. In this section, we discuss a series of tests and comparisons that are useful in diagnostic analysis of the photochemical model. The general categories of these analyses include: sensitivity/uncertainty testing, testing of species other than ozone (when adequate data permit), examination of model-predicted fluxes and pollutant budgets, and the analysis of residuals. In some cases the tests we recommend must be supported by high-quality aerometric data bases, only now available in some areas. As modeling techniques mature and as data bases improve, the techniques in this section may well become part of the routine model performance evaluation process in the next several years.

3.1 Sensitivity-Uncertainty Analysis

Sensitivity analysis consists of systematically studying the behavior of a model over ranges in variation of inputs and parameters. This process may actually extend to studying the behavior of the model for changes in its basic structure, i.e., for different assumptions in its formulation. When model inputs and parameters are varied over their ranges of uncertainty in order to provide estimates of the uncertainty in predicted concentrations due to these input uncertainties, the process can be termed a sensitivity-uncertainty analysis. The diagnostic simulations outlined in Section 2.3 fall within the general category of sensitivity analysis. We presented them in Section 2.0 as they

represent operational tests that should accompany any model performance evaluation.

Sensitivity analysis can be used to determine if the predictive behavior of a model is consistent with that expected on the basis of its underlying chemistry and physics, namely does the model respond "properly" when its inputs and parameters are varied. Sensitivity-uncertainty analysis, just a sensitivity analysis, in which the variations in inputs and parameters correspond to their estimated uncertainties is used to estimate the uncertainty in a model prediction, and may be viewed as a form of error analysis (Munn et al., 1988). In short, sensitivity analysis of air quality models serves to meet two objectives: (1) to determine in a qualitative sense if a model responds to changes in its inputs and parameters in a manner consistent with what is understood about its basis physics and chemistry, and (2) to quantitatively estimate the uncertainty in model predictions arising from uncertainties in the inputs and parameters.

Various methods applicable to sensitivity-uncertainty analysis of photochemical air quality models are discussed in the literature (McRae et al., 1982; Dunker, 1980, 1984, 1986; Tesche et al., 1981; Tilden et al., 1981; Seigneur et al., 1981; Brost, 1988; and Derwent and Hov, 1988). An overview and synopsis of major results of sensitivity testing and analyses of photochemical air quality models can be found in Seinfeld (1988a).

Parameters of interest in the sensitivity analysis of photochemical air quality models include:

- Structural and design parameters of the model. Examples are the horizontal and vertical dimensions of the computational grid cell, the number of cell layers in the vertical direction, the size of integration time steps, etc. Changes in these parameters are deliberate and related to model use. The objective of sensitivity testing here is to identify values of such parameters leading to an optimal combination of computational efficiency and accuracy of predictions,

- Constitutive parameters of the model. Examples are chemical reaction rate constants, deposition velocities, etc. Sensitivity analysis usually focuses on the effects of uncertainty in the values of these parameters on model predictions,
- Input parameters. These are calculated from the input data and, as discussed below, carry the uncertainties inherent in these data.

The influence of the data input to the model on its predictions may also be studied through sensitivity analysis. These data are used to drive the model either directly or through the determination of input parameters by "preprocessors" or "peripheral models." Whether a component of the model that describes atmospheric processes--such as, for example, the windfield generator--is considered a preprocessor or a component of the main photochemical model is mostly a matter of programming style and computational efficiency. Models unrelated to atmospheric processes, such as emission models, are typically considered peripheral models. Input data include uncertainties due to both measurement errors and to the randomness of the atmosphere. Uncertainties are also introduced by various assumptions and approximations in data processing.

Sensitivity analysis with respect to a single parameter or input that affects one component (process module) of the model. Alternatively, sensitivity testing can be done simultaneously with respect to more than one parameter (or inputs), that, however still affect only one component (process module) of the model (e.g., the chemical reaction rate coefficients). Related is sensitivity testing with respect to a single parameter that affects more than one (interacting) process in the model. For example, changing the resolution of the model affects the interaction of transport/mixing and chemistry. Whether sensitivity analysis is concerned with changes in a parameter that affect one or more interacting processes depends on the study objective. For example, one can examine the sensitivity of a model to the windfield inputs or vertical dispersion parameters by considering conserved tracers. Alternatively, one can study the effect of changes in the windfield, or dispersion parameters on the overall evolution of the photochemical system. Usually both approaches are

needed in order to provide sufficient understanding of the system dynamics. Finally, multiparametric sensitivity testing can be carried out for a group of parameters (or inputs) that affect more than one (interacting) processes in the model.

Sensitivity analysis of the first two types mentioned above would not require running the full photochemical model; rather, only the specific module (chemical mechanism, windfield preprocessor, etc.) that is affected by the parameter changes would need to be examined. In some cases it might also be possible to perform the latter two types of sensitivity tests by employing a simplified model (see Section 4.5), but the limitations inherent in such an approach should not be overlooked.

3.1.1 Sensitivity-Uncertainty Analysis

When the input data or parameters to be studied have been identified, a sensitivity-uncertainty analysis can be carried out. Ideally the probabilistic distribution of the input parameters, incorporating random variations due to both natural variability (from, e.g., emissions, meteorology) and "errors" are available. Errors include both measurement uncertainties and, in the cases where the input parameter is derived by modeling of "raw" or "basic" input data, such as mobile source emissions, uncertainties related to the assumptions involved in the modeling step. In practice, one usually must settle for much less information on input uncertainty than the complete probability distribution. In fact, model response can often be rather insensitive to the exact shape of the input distribution (O'Neill et al., 1982; Alcamo and Bartnicki, 1987). Typically, only estimates of the range of uncertainty are available. Sensitivity-uncertainty analysis then consists of determining the outputs of the model resulting from inputs prescribed at the limits of this range.

The sensitivity-uncertainty analysis involves determining how the uncertainties in input data and parameters propagate through the model equations, producing uncertainties in the output. Most techniques for uncertainty propagation through model equations are based on Monte Carlo methods. These methods involve random sampling from the distribution of inputs with

corresponding model runs to obtain a statistically significant distribution of model outputs. Modified Monte Carlo methods perform the sampling from the input distribution in an efficient ("almost optimal") manner so that the number of necessary solutions, compared to the brute force Monte Carlo method is greatly reduced. Three such methods appropriate for use with air quality models are the Fourier Amplitude Sensitivity Test (FAST) (McRae et al., 1982), Latin Hypercube Sampling (Iman et al., 1980; Derwent and Hov, 1988), and Fractional Factorial Designs (Teschke et al., 1976).

As noted, sensitivity analysis has the objectives of: (1) evaluating the behavior of the model with respect to changes in inputs and parameters, and (2) providing quantitative estimates of the uncertainty associated with model predictions. A sensitivity analysis should normally accompany a major model performance evaluation study. The key input variables and parameters need to be identified and their range of uncertainty estimated. These uncertainty bounds are then propagated through the model, either singly or in concert, to provide estimates of the uncertainties in the predictions. The modeler should discuss the results of the sensitivity runs from the viewpoint of the response of the model expected from its underlying analysis and chemistry. Typical inputs and parameters that should be considered for sensitivity analysis include:

- Key emission parameters, such as VOC rates and speciation;
- Mixing height;
- Wind field; and
- Deposition velocities.

A critical issue driving the entire question of model performance evaluation is what effect do uncertainties in model predictions have on emission reduction estimates obtained from a model. For example, if the ROG emissions inventory is judged to have an uncertainty of ± 50 ppb, what is the effect of this uncertainty when ROG and NO_x emissions reduction scenarios are evaluated? If a model is to be used for emission reduction calculations then it is useful to perform an appropriate sensitivity analysis.

3.1.2 Characterization of Uncertainties and Comparisons of Uncertain Predictions and Observations

Photochemical model predictions, although often stated and interpreted in a deterministic framework, always include reducible and irreducible uncertainty in both input data and parameters and in model formulation. Sensitivity studies may provide quantitative estimates of these uncertainties.

When comparing model predictions and observations, one must remember that observations contain uncertainties due to measurement errors and the naturally random character of the atmosphere. Fox (1981) gives a good introduction to the concept of uncertainty in air quality modeling. Beck (1987) offers further discussion of the concept of uncertainty in environmental models and data.

Ideally, comparisons of predictions and observations should account for uncertainties in both the observational data and the model predictions. One then seeks to determine if the uncertainty envelope around the data and the uncertainty envelope around the predictions overlap sufficiently. Figure 2-17 illustrates this concept. Error basis on the observational data form the outline of an "observation envelope," while the "prediction envelope" has been generated by varying the mixing height over a presumed range of uncertainty. This section is devoted to a discussion of how the comparison of observation and prediction envelopes could be carried out in a quantitative manner. While such comparisons are not yet routinely carried out as a part of the model performance evaluation process, this type of comparison may be quite valuable when comparisons of predictions and observations in the absence of explicit accounting for uncertainties, such as discussed in Section 2.0, yield comparison statistics that are poorer than one would seek for a verified model.

Simple descriptive statistics such as the numerical performance measures suggested in Section 2.0 help summarize the results of the evaluation process, however, these statistics do not account for uncertainties in either the data or the predictions. Interpretation of numerical performance measures such as bias and error is enhanced by the estimation of certain statistical tests

such as confidence intervals. Indeed, the 1980 AMS/EPA workshop on judging air quality model performance (Fox, 1981) recommended that evaluation of performance measures should be stated in the form of confidence intervals. Specifically, Fox states "Statistical tests (Student's t or Wilcoxon) could be used to indicate whether the average observed concentration and the average predicted concentration for a particular meteorological category were significantly different from one another. An F-test could be used to compare the estimated variances of the observed and predicted concentrations for each category." He also notes that "It is reasonable to urge that the use of statistical confidence interval statements be accepted as a common tool in model performance evaluation and as a recommended initial step towards setting performance standards. A confidence interval statement for a parameter is more informative than a statistical test. The interval statement provides information relative to many different hypotheses, whereas the statistical acceptance/rejection test is a decision rule about a single hypothesis... Interval statements can be constructed for the bias and the variance using the Student's t and the chi-square distribution."

Meaningful use of statistical tests and confidence intervals requires a random sample. The problem here is that data needs for a systematic diagnosis of errors in the model may conflict with the need for a random sample. For example, usually days having particular meteorological conditions relevant to high ozone levels are selected for modeling. Although this is the sample of interest it cannot be used as a basis for statistical inference because there is no way of estimating the bias introduced by the selection of that particular set of days. In addition, the sample size must be "large enough". For a given population, a larger sample produces more precise estimates of statistical measures and, hence, narrower confidence intervals. It is important, therefore, to obtain data and simulate as many days as possible.

3.2 Analysis of Residuals

A residual is the difference between a prediction and its corresponding observation, the remnant in variability after the systematic or ordered component of the observation, as explained by the model, is removed. The set

of residuals generated in a typical performance evaluation exercise contains all available information about contributions to "error" of (a) the concentration data used for comparison with model output, (b) the soundness of the model formulation, and (c) the adequacy of the data supplied as input to the model. If the model describes actual dynamic behavior reasonably well, and if the observational data are reasonably accurate and representative, then the residuals should have the characteristics of noise--little or no bias and a standard (typically, normal or lognormal) distribution of error, with moderate to small variance. If, on the other hand, the residuals display bias or other systematic characteristics, the model is not fully or correctly representing the actual system (assuming that the data being used in the comparison are indeed accurate and representative).

When unexplained but statistically significant variation characterizes the residuals, and concerns about model inadequacy arise, plotting residuals against selected variables is an attractive means for identifying patterns of aberrant behavior. If correlations (relationships) between the residuals and one or more selected variables can be found, the emergent pattern (s) may be suggestive of the cause(s) of failure or inadequacy in the model. If the cause(s) of model inadequacy can indeed be found and corrected and the residuals replotted, the correlation(s) previously seen, in principle, should be reduced or eliminated. The characteristics of the new residuals presumably approach those of noise.

A wide range in variables may be selected for plotting against residuals, including:

- Time;

- Geographical location;
 - Along the general direction of the wind, and
 - Overall spatial pattern (two dimensions).

- Concentration levels;
 - Precursor (hydrocarbon and NO_x, for ozone) concentrations,
 - Concentrations at the inflow boundary (as they change with time), and
 - Estimated concentrations of intermediate species, such as free radicals.

- Meteorological variables (and their spatial and/or temporal patterns);
 - Wind speed,
 - Wind direction,
 - Ventilation,
 - Height of the mixed layer, and
 - Solar insolation.

- Emissions (as they change with time);
 - Activity levels, and
 - Rates.

- Deposition rates (as they change with time);
 - Activity levels, and
 - Rates.

Such plots can be made (or correlations calculated) for the full region of interest and for the full duration of the simulation, or for subregions, selected time periods, and specified ranges in variables. This list of variables and the subsets of conditions suggested are by no means exhaust the possibilities. Hypotheses that are advanced concerning potential flaws in a model should suggest variables and circumstances for investigations. Diagnostic analysis is an art, enhanced by the knowledge and inquisitiveness of the investigator. Seeking explanations and finding solutions is central to the methods of science; analysis of residuals is one of a number of methods available.

3.3 Multi-Species Comparisons

The development of evaluation procedures that test photochemical model performance for species other than ozone has been strongly recommended by the Urban Airshed Model Technical Advisory Committee (Seinfeld et al., 1987) and the SCAQS Model Working Group and Data Analysis Group (Seinfeld et al., 1987, 1988). The reasoning for these recommendations is that multi-species comparisons can provide a more robust basis for accepting or rejecting a model (or a model simulation); they significantly improve the chances that a flawed model will be identified. Adequate model performance for several reactive species increases the decision-maker's assurance that correct ozone predictions are not a result of chance or fortuitous cancellation of errors introduced by various assumptions. Multispecies comparisons may be the key in discriminating among alternative modeling approaches that provide similar predictions of ozone levels.

To date, only limited comparisons have been published on model performance for species other than ozone in photochemical simulations. Most model evaluation studies available present only ozone results (Tesche, 1988) although there are a few limited tabulations of NO₂ predictions (Wagner and Ranzieri, 1984). Studies reported by Roth et al. (1983), Tesche et al. (1983), Russell and Cass (1986), Wagner and Croes (1986), and Russell, McCue and Cass (1988) are among the few that present performance evaluation statistics for associated pollutants. (Lack of ambient measurements for such pollutants is the major reason for the limited number of past studies.) The SCAQS and other recent data bases (e.g. San Diego, Sacramento, and San Francisco) now offer the potential for several comparisons with species other than ozone. The SCAQS data in particular afford a level of testing of photochemical models and modules, such as the chemistry mechanism, not previously possible. The availability of ambient air measurements for speciated organics, key species such as HCHO, PAN, NO₂, H₂O₂, and HNO₃, and organic acids will allow not only more extensive "operational" model testing but also diagnostic and comparative evaluations. Finally, the SCAQS data base, or similar data bases that will be assembled in the future (e.g., the San Joaquin Valley Air Quality Study), also offer the potential for mechanistic evaluations of alternative chemical kinetic mechanisms.

Until recently, such evaluations have been based entirely on smog chamber data (see Section 4.2.4).

Evaluation of model performance for precursor and intermediate species as well as for product species other than ozone is recommended when ambient concentration data for these species are available. Comparisons of observed and predicted concentration values for all important chemical compounds involved in photochemical air pollution, inorganic and organic precursors, intermediates, and products such as O_3 , NO, NO_2 , PAN, individual hydrocarbons, H_2O_2 , HONO, and HNO_3 , is useful in model evaluation, especially with respect to the chemistry component of the model. As discussed in more detail in Section 4.6, comparisons of predictions and observations for total organic nitrates (mainly PAN) and inorganic nitrates (HNO_3 and nitrate aerosol) can be used to test qualitatively if the emissions inventory has the correct relative amounts of ROG and NO_x . However, HNO_3 and nitrate aerosol cannot be included in the data set for model comparisons if the model does not include an adequate description of the HNO_3 depletion process associated with aerosol formation.

In some cases, modifications to existing models or data handling procedures will be required in order to permit the types of multi-species comparisons we recommend. For example, at the present time, non-reactive hydrocarbons are not input into photochemical models. Typically, the non-reactive nonmethane compounds comprise 20 percent to 30 percent of the nonmethane carbons. Another model modification that would reduce ambiguity in performance evaluation is to include non-reactive hydrocarbons as a transported species. This would allow more direct comparison of the VOC predictions and ambient data.

3.3.1 Selection and Organization of Predicted/Observed Data

Depending on the objectives of the performance evaluation, multi-species comparisons can be organized in various ways, depending on the number and kind of chemical species selected. For example, in applications involving long-range transport, where multiday photochemistry is important and less reactive organics play a significant role, one could group and compare

species according to their average atmospheric lifetimes. In general, however, the selection of the species should reflect their importance and role in determining ozone dynamics.

As a general recommendation, the following species have the highest priority for prediction-observation comparisons:

- Source/sink species (NO, NO₂, VOC speciated). Reliably reproducing NO and NO₂ concentrations should build confidence that the model treats the nitrogen balance correctly. However, the ability of a model to accurately reproduce individual VOC observations is in general limited; relevant problems are discussed later.
- Intermediate species (H₂O₂, HCHO, PAN). Comparisons of H₂O₂ and HCHO predictions and observations will allow one to evaluate the representation of the radical chemistry in the model. Satisfactory prediction of PAN should suggest that one key radical sink is treated correctly.

Depending on the availability of other measurements and the incorporation of aerosol dynamics and thermodynamic processes in the model, the above comparisons can be supplemented with others. For example, because HNO₃ is a sink for the OH radical and because H₂O₂ is a sink for the HO₂ radical, the ratio of HNO₃ to H₂O₂ is an indicator of the extent to which OH and HO₂ radicals are adequately simulated.

3.3.2 Performance Measures and Statistical Tests

Selection of numerical and graphical performance measures for multi-species comparisons should follow methodology outlined in Section 2.0 for comparing observed and predicted ozone concentrations. One may also wish to compare distributions of predictions and observations and using quantile-quantile graphs and statistical tests, if adequate data for such comparisons are available. Because of the large uncertainty of measurements at low concentration

levels, comparisons of observations and predictions should not be made with measurements near the detection limits of the measuring instruments.

3.3.3 Graphical Procedures

Graphical time series plots of selected concentration ratios of chemically linked species (e.g. NO_2/NO , PAN/NO_x , $\text{HCHO}/\text{H}_2\text{O}_2/\text{O}_3$) may facilitate the comparison of predictions and observations or of predictions from different models. Barchet (1987) suggests that these ratios may be the preferred indicators of gas-phase chemistry module performance because they are less sensitive to non-chemical factors such as transport, dispersion, and dry deposition. This may not always be true, depending on the relative magnitudes of the characteristic time scales of chemical and transport phenomena. For example the NO_2/NO ratio in the vicinity of large localized sources is governed by the state of mixing rather than the chemistry. However, the use of ratios will indicate prevailing trends more clearly. Disagreement of predicted and observed concentration ratios where the gas-phase chemistry module has shown satisfactory performance may indicate the need for a more comprehensive treatment of non-chemical processes. It has been suggested that analyses of ratios of chemical constituents having different chemical lifetimes can be used to provide a basis for differentiating removal through physical processes versus removal through chemical processes (EPA, 1988).

3.3.4 Measurements and Measurement Evaluation

In using and interpreting (or when planning to measure and use) observed ambient concentration levels of the groups of species selected for a multi-species model performance evaluation, it is important to keep in mind that such measurements are non-routine, that the associated uncertainties may be significant, and that, most probably, the available number of data points for a given modeling episode will be limited. Hence, one should be cautious when drawing conclusions regarding model performance when non-routine measurements are used.

The location of the monitoring sites for field measurements of the additional species should reflect the intended use of the predicted and observed concentration comparisons in the performance evaluation process. Total VOC should be measured during well mixed conditions at downwind locations to match the emission flux. Speciated VOC should be measured during morning hours near sources (ideally both at the surface and aloft) to check emission inventory speciation. PAN, HCHO, and H₂O₂ should be measured in the general area of the ozone peaks to provide information needed to check the fidelity of the chemical mechanism. General guidance for the location of monitoring stations for traditionally measured photochemical pollutants can be found in Ludwig and Shelar (1978).

There are practical limitations in evaluating a model's performance in predicting speciated VOC. Conceptually, comparisons can be made between observed and predicted total VOC concentrations as well as between observed and predicted concentrations of classes of VOC. The second type of comparisons requires aggregations of the ambient VOC into the classes used in the particular chemical mechanism employed by the model under consideration. However, emissions of non-reactive organic compounds, which may comprise 5 percent to 30 percent of the actual ROG emissions, usually are not included in the simulation. Thus, predicted VOC may have an inherent bias toward underestimation relative to observed VOC unless the difference is accounted for by explicitly excluding the non-reactive compounds from the observed concentrations. Also, fairly severe approximations are made in the VOC lumping schemes. In some cases, the assignment of individual compounds to a class is based more on the similarity of their ozone formation potential to that of the model species rather than on the rate at which they react. Hence, perfect agreement is not expected and, in fact, agreement within ± 20 percent for VOC classes is probably the best one can expect with current lumping schemes.

Another method of comparing model predictions with measured ambient concentrations involves the use of atmospheric tracer experiments. Experiments that involve inert materials (SF₆, freon, or perfluorocarbons) released at known locations and in precise amounts provide direct observational evidence of transport and dispersion processes and afford an opportunity for independently

evaluating the relevant components (transport and dispersion modules) of the grid-based photochemical model. The methodologies for operational and diagnostic testing of grid-based models using experimental inert tracer data are well established (Tesche, Haney, and Morris, 1987). A discussion of meteorological model evaluation, often considerably enhanced by availability of tracer data, is presented in Section 4.2.2.

In sum, multi-species comparisons are needed to evaluate the performance of a photochemical model in predicting ambient levels of other criteria pollutants such as NO_2 , and to establish that the model predictions of the temporal and spatial ozone dynamics are "correct for the right reasons" and hence to discriminate among photochemical models that predict similar ozone concentrations. Priority for multi-species comparisons should be given to NO , NO_2 , speciated VOC, as well as to PAN, HCHO, and H_2O_2 . The numerical and graphical procedures outlined in Section 2.0 may be used for this purpose. Appropriate ratios of the concentrations of these species (observed and predicted) are also recommended in this diagnostic evaluation process.

3.3.5 Model Modifications to Facilitate Testing

One of the most difficult problems in diagnostic evaluation of photochemical models is confirming that the emissions inventory has the correct relative amounts of ROG and NO_x . It is difficult to confirm this by direct comparison of observed and predicted precursor concentrations because of the incommensurability problem. Another method of assessing whether the relative amounts of ROG and NO_x are correct is to evaluate the relative amounts of organic and inorganic nitrate formed by the model. Based on the current understanding of the atmospheric chemistry, biases in the ratio of organic nitrate to total nitrate should qualitatively reflect biases in the ROG to NO_x ratio of the emissions.

Historically, data for the various nitrates either were not available or were highly uncertain due to measurement problems. Improved techniques are now available and have been used in measure nitric acid and aerosol nitrate as well as PAN in numerous field programs. Few, if any, measurements exist for

organic nitrates other than PAN; however, we expect that these "other organic nitrates" are a small part of the nitrogenous products. Typically, the condensed chemical mechanisms predict the concentrations of PANs, the sum of PAN and PAN analogs (usually labelled as PAN) and total inorganic nitrate (TIN), which is the sum of nitric acid and aerosol nitrate (usually labelled HNO₃). Some mechanisms may also predict the concentrations of other organic nitrates.

The inorganic nitrate predictions from contemporary photochemical models are probably fairly inaccurate because they lump together two species that deposit at very different rates (i.e., nitric acid deposits very rapidly, while aerosol nitrate deposits slowly). However, if the models are modified to incorporate the nitric acid-aerosol nitrate equilibrium chemistry, as in the models developed by Hogo et al. (1985), Russell and Cass (1986), and Russell et al. (1988), then their nitric acid and aerosol nitrate predictions can become useful components of the multi-species comparison. Comparison of the observed and predicted ratios of PAN to total nitrate and TIN to total nitrate should be made to assess possible biases in the VOC and NO_x inputs to the models.

3.4 Mass Fluxes and Budgets

Only recently have attempts been made to derive mass balances and carry out flux calculations for photochemical grid model simulations. This has occurred more routinely for regional-scale Eulerian models. Four mass balance and flux calculation procedures are suggested to accompany detailed performance evaluations. The first procedure involves computing the mass fluxes into and out of the domain boundaries. The fluxes of individual precursor species (NO, SO₂, CO, organics) and product species (NO₂, O₃, HNO₃, SO₄) into and out of the four side walls and the top of the modeling domain should be calculated each hour and summed for each simulation day. Hourly and daily flux totals should be reported. These calculations could be performed so that concentration isopleths of the species distributions crossing the four side walls and the top of the modeling region could be displayed for further diagnostic analysis.

The second procedure involves the mass fluxes into and out of the mixed layer. Here, hourly and daily average mass fluxes of all transported species

should be calculated for vertical transfer into and out of the mixed layer. These flux arrays may also be contoured for diagnostic analysis. Third, the surface deposition fluxes should be estimated. Hourly and daily average surface deposition rates should be calculated and reported for each species removed at the ground. This information should be presented as basinwide hourly and daily totals and by means of a daily surface deposition isopleth.

The final procedure involves the reconciliation of emissions, transport, transformation, and removal terms in a simplified, closed mass budget over the whole modeling domain. The various flux terms described above, when combined with the hourly emissions rates, may be used in a simple mass budget to apportion the total mass in the modeling domain into emission, transport, and removal components. The transformation term is obtained by difference, assuming a closed budget. The degree of imbalance in the hourly and daily mass budget should be reported as well. Mass balances and flux calculations have been performed only to a very limited extent in the past and there is little present guidance that can be offered with respect to how these results should be judged. The true value of these calculations for diagnostic performance evaluation and stress testing will evolve as more experience is gained in their use and interpretation.

4.0 RESEARCH NEEDS IN MODEL PERFORMANCE EVALUATION

Many of the key problems in photochemical model performance evaluation have been known for a long time. Those amenable to near-term solution, thorough either emerging high-quality data bases (e.g., SCAQS, SJVAQS) or added emphasis given to the evaluation process, were discussed in earlier sections. Over the longer term, there are several problems whose solution will come only through more research, acquisition of specialized data bases, and the commitment of significantly greater resources to the task of model evaluation than are available today. In this section, we discuss these longer-term research needs.

4.1 Testing the Adequacy of Model Response to Emission Changes

Assessing the ability of models to correctly simulate the effects of emission changes represents a major need in model performance evaluation since it is directly related to the intended regulatory application of photochemical models. Traditionally such models are evaluated for a variety of meteorological conditions over periods of time that do not involve major changes in emissions. Then, the critical assumption is implicitly made that the models would also perform in a satisfactory manner under conditions of drastically altered emissions. However, there is no evidence to support this position. The work of Dennis and co-workers (Dennis et al., 1983; Dennis and Downton, 1984; Downton and Dennis, 1985; Dennis, 1986) has shown that grid-based photochemical models which perform adequately over a range of meteorological conditions do not necessarily show similar results when the evaluation involves a large change in emissions. They found that different versions of the Urban Airshed Model (UAM), which gave similar performance results under conditions of changing meteorology, performed very differently when tested for emissions changes. (These versions of the UAM represented progressive improvements in chemistry, numerical methods, and the treatment of meteorology.) It is imperative, therefore, to evaluate photochemical models that are intended for use in the development of air pollution control strategies to determine their ability to simulate the effects of emissions changes.

The problems raised in this type of evaluation are serious. Detailed emission inventories of comparable accuracy are required for years sufficiently far apart (e.g., 10 years) so that major emission changes have taken place to allow a meaningful model performance evaluation. One has also to identify episodes occurring in periods of similar meteorology; the importance of meteorology in modeling emission changes is discussed further later in this section. Ideally, this type of evaluation involving historical emission and air quality records is the most preferable one, but generally the lack of detailed inventories and historical aerometric data prohibits this approach. Even if the required data are available, one would have to take into account all changes in the procedures for developing emission inventories, for monitoring air quality, and so on that have occurred over the years. So, almost unavoidably, stringent assumptions and approximations regarding past emissions and ambient air quality (initial and boundary conditions and ozone levels) will have to be made when modeling episodes from past years. This problem may diminish somewhat in the future, at least for some areas. The use of weekday versus weekend emission rates has been suggested as an alternative to retrospective modeling. Even assuming the inventories are accurately estimated, it is doubtful that the level of emission change would be sufficient for a meaningful model performance evaluation, although such studies would no doubt be valuable. A third promising approach in evaluating a model's ability to correctly predict the effects of major changes in emissions is to thoroughly test the photochemical model for different urban areas, using input data sets of similar quality.

Ideally, evaluation of the model (for a range of meteorological conditions) should be performed using emission inventories and aerometric input data for the same region and for years sufficiently far apart. Comparable quality of inputs (mostly emission inventories) should be assured for such a task.

Since evaluation using historical inventories and aerometric data does not appear feasible at the present time, the alternative of evaluating the model on the basis of its performance for different urban areas should be adopted. A fixed version of the photochemical model (same horizontal and vertical resolution, identical input data preprocessors, chemistry and removal modules, and so on) should be applied to all regions selected for the evaluation. The

evaluation should span a wide range of meteorological conditions for the urban areas under consideration, corresponding to high, moderate, and low ozone days. Such an evaluation does not test uniform changes in emission levels. Instead, overall model performance for different spatial and temporal distributions and source strengths, and speciation of emissions is provided.

Even if the problems of availability and quality of input data (emission and aerometric) are solved, allowing one to evaluate a model's ability to simulate significant emissions changes, one must still account for the fact that the sensitivity of a photochemical modeling simulation to emission changes will vary according to meteorology. Wagner and Wheeler (1989), reporting on sets of simulations performed by Tesche et al. (1988a, 1988b) and Wagner (1988) concluded that: "The location and amount of maximum sensitivity to emissions changes vary with the meteorology. This may mean that more than one episode should be used in evaluating the effects of emission changes upon peak ozone concentrations." Indeed, the selection of particular ozone episodes on which to design emission controls can have a substantial effect on the projected control levels. It is important to examine several episodes to determine the sensitivity of control levels to meteorology.

Considerable effort has been devoted in recent years to developing "screening/regression" methods that decouple the effects of meteorology from those of emission changes on observed high ozone concentrations (Horie, 1987; Pollack et al., 1988; Stoeckenius, 1989). This technique is commonly referred to as "meteorological de-trending" of an air pollution episode. Further testing and research into methods for identifying an adequate set of meteorological episodes that may be used in meteorological de-trending analyses is needed.

In sum, the adequacy of a photochemical model in correctly predicting the effects of emission changes on ambient air quality should be evaluated directly, by examining model performance for applications involving significantly altered emission strengths and spatial patterns. Ideally such a performance evaluation should be performed with emission inventories for the same region that correspond to years sufficiently apart. When such a procedure is not feasible, one should evaluate the model for two or more different regions, using

input data of comparable quality. The effects of meteorology on the model's response to emission changes must be carefully analyzed.

4.2 Mechanistic Model Evaluations

A mechanistic model evaluation is an assessment of the ability of an individual process modules' ability to reproduce the observed salient features of the process. When applied to all process modules that constitute the full photochemical dispersion model, it represents a test of the correctness of the underlying science. In this section, we recommend mechanistic evaluation procedures for four major process modules used to support photochemical modeling: emissions, meteorology, chemistry, and deposition.

4.2.1 Emissions Model Testing

The emissions inventory input to a photochemical model is itself a model in the same sense that the windfield generation routine and the chemical reaction mechanism are models. Whereas the uncertainties in meteorological and chemical mechanism models have received extensive attention, those associated with the emissions inventory generally have received far less consideration in the modeling context. This is due, no doubt, to the great difficulty in quantifying emissions uncertainties. Even though much effort has been expended to assemble adequate emissions inventories, there is significant feeling that the emissions inventory may be the least accurate input in the photochemical modeling process.

Estimation of uncertainties associated with emissions inventories is of considerable importance since any control strategy decisions made, with or without the aid of models, are only as good as the emissions estimates upon which they are based. Subjective "Delphi" estimates of emissions inventory uncertainties are of little use and should be replaced with other, improved approaches (see, for example, Benkovitz and Oden, 1989; Dickson and Hobbs, 1989). Some new directions in the evaluation of emission inventory models are suggested below.

"Top down" vs. "Bottom up" Calculations

The emissions inventory should be developed for a pollutant by estimating emissions of individual sources or groups of sources and then aggregating them and, independently, by calculating gross or integrated emissions rates. Where possible, one should compare the two estimates and use the information gained from analysis of differences in the diagnosis of sources and magnitudes of uncertainty. Generic procedures for this "top down-bottom up" approach need to be developed and potential areas of applicability of this approach, including routine or special inventory development and special studies to assess uncertainties, need to be specified.

Mass Balances

There is a need to develop means for carrying out mass balance calculations as part of the inventorying process. It is necessary to prescribe data needs, anticipated uncertainties in determination of individual elements, calculational procedures, and the like. For example, nitrogen balances can be made for determining emissions estimates and uncertainties for animal wastes, chemical fertilizer use, wastewater treatment plant emissions, and so on.

Ambient Air Ratios vs. Emissions Ratios

Another method of evaluating the emission inventory is to compare ratios of ambient concentrations of two selected pollutant species, measured near the source, with estimated ratios of emissions rates of the two species. Information gained from analysis of differences in ratios may be used in estimating the magnitudes of emissions uncertainties (and in suggesting means for reducing uncertainties). As examples, the ratios of toluene/xylene and toluene/paraffin might be used in such comparisons.

Indirect Confirmations

In assessing the emissions inventory, one should use external information to the fullest extent possible to corroborate direct emissions estimates.

One recent example involves the iterative use of an energy balance to reduce uncertainties in fuel consumption statistics, which in turn are used in estimating sulfur emission rates. Other consistency checks include: (a) use of basin-wide box models for long averaging times, as for lead, (b) comparison of spatial distributions of emissions and air quality, as for ammonia in the SOGAB, (c) comparison of relative magnitudes of mass in the atmosphere and in the inventory, as for trace metals, and (d) examination of photochemical modeling calculations within a particular air basin to infer if there is a tendency for pronounced bias in various components of the inventory.

Source Testing

One should determine emissions of selected sources through direct measurement. i.e., source testing. Employing statistical sampling procedures for selecting a subset of a large number of sources in a category for testing may be quite valuable. Generally speaking, if knowledge of emissions is inadequate, one should explore the feasibility of measuring emissions in the field. Because of the large attendant costs, one must develop specific guidelines for establishing this need and procedures for cost-benefit assessment in order to justify this major effort.

Designed Field Studies

Thoughtfully-designed field studies should be considered specifically for the purpose of emissions determination. Examples include the CRC-sponsored SCAQS-1 "Tunnel Study" for corroborating estimates of emissions from vehicles and (Ingalls et al., 1989) and the recent CO determinations at roadways in Denver and in Lynnwood, CA. These field studies may include the release of inert tracers from sources of interest, monitoring their concentrations at downwind locations. Rare-earth doping of fuels in European experiments is another example.

General Guidelines for Inventories

One should determine, to the extent possible, the accuracy required to satisfy the needs of policy makers. This leads to sensitivity studies to

determine the accuracy and precision necessary for individual components of an inventory. Apart from traditional simulations, these investigations might include calculating sensitivities associated with the temporal distribution and modal split in the driving cycle to determine importance of developing a weekend inventory. For ozone modeling, it is highly desirable to determine the model's sensitivity to accuracy of the aldehyde inventory estimates and to the emissions factors used to develop on-road motor vehicle emissions.

Emission inventory specialists, modelers and decision-makers should determine and be cognizant of realistic limits on achievable accuracy. The inventory should be neither too comprehensive nor too "lumped." Proper spatial, temporal, and chemical resolution must be achieved, realizing that these needs must be balanced by other demands for emission inventory information.

There is a clear need to improve the characterization of the driving cycle, including the degree of representativeness, extent of variability due to changes in commute characteristics, driver characteristics, and other influences. Also, there is a need to develop better approaches to estimating spatial and temporal uncertainties associated with grid-based inventories. With respect to the entire inventory development process, there is a crucial need to provide appropriately detailed documentation of the inventory development process. This includes the preparation of documented source code that processes the various basic emissions data and parameter files into gridded, temporally-resolved, and chemically-speciated photochemical model inputs. The Emissions Preprocessor System (EPS) recently developed for the UAM (SAI, 1989) is an example of such an "emissions model".

4.2.2 Meteorological Model Testing

A wide range of meteorological models are available to support grid-based photochemical modeling. Meteorological models provide estimates of the two- or three- dimensional wind patterns, surface and aloft temperature fields, estimates of the mixing height field, or all of these variables, depending on the model selected. Moreover, the domain scales over which the model may be applied may range from local to regional. In this discussion, we focus on

diagnostic and prognostic meteorological models that estimate winds, temperatures, and mixing heights over mesoscale domains because this is the relevant size of urban- and regional-scale photochemical modeling grids. We do not consider the objective analysis class of meteorological models (e.g., Goodin, McRae, and Seinfeld, 1980) due to the inherent problems that arise when attempting to evaluate a model with the SCN data used to operate it.

Complex meteorological models are not necessarily superior to simpler models for supplying inputs to photochemical models in a particular application. If, for example, the observational network is exceptionally dense, objective analysis methods (i.e., interpolation techniques) or simple diagnostic models may be cost-effective alternatives to the more complicated "primitive equation" models. But, advanced numerical meteorological models are expected to provide greater overall consistency in their predictions because they contain more of the relevant atmospheric physics. A brief review of previous meteorological performance evaluation procedures is presented below, followed by a recommended set of numerical and graphical procedures for mechanistic evaluation of meteorological models.

Previous Model Evaluation Procedures

Since the mid-1970s, there has been considerable work in developing procedures for evaluating air quality model performance. Less work has been performed in developing procedures and standards for meteorological models used in support air quality modeling. While there is a rich history of performance evaluation of meteorological models used in weather forecasting (e.g., Anthes, 1983; Ray, 1986), these studies are of limited use in photochemical modeling for three reasons. First, the time and space scales relevant to ozone modeling are characteristically shorter than those corresponding to the weather prediction problem. The regional-scale (or limited area) models have typical horizontal resolutions of 50-250 km, compared with the photochemical model resolutions of order 2 to 5 km. Second, evaluation procedures for limited-area weather prediction models have generally consisted of "skill" measures (related to the calculation of large-scale pressure gradient fields), root-mean-square errors, and correlations between forecasted and observed changes in surface fields of

wind speed, temperature, pressure, or precipitation. These procedures have not adequately stressed the model or data bases used to operate them.

Third, weather prediction models forecast some future atmospheric state, given an initial characterization. Meteorological modeling for air quality problems, in contrast, is diagnostic in nature. One is principally interested in reproducing the local meteorological details of some past ozone episode that typically lasted 24-72 hours. Given an initial description of the atmosphere at the beginning of the episode and meteorological data throughout the period for supplying boundary conditions, models are used to reproduce the local wind, temperature, and stability fields at all grid points within the computational domain. Although essentially a diagnostic problem, both steady-state and time dependent (i.e. prognostic) models are used to re-create these past meteorological fields.

Qualitative Evaluations

Numerous qualitative evaluations of mesoscale meteorological models appear in the literature (Fosberg et al., 1976; Martin, 1981; McNider, 1981; Segal et al., 1982a,b; Garrett and Smith, 1984; Pielke, 1984, 1985; Kessler and Douglas, 1989; Yamada et al., 1989). Generally, these studies compare gridded surface wind patterns predicted by a model with those developed from streamline analysis or by simple interpolation of observational data. Vertical profiles of temperature and wind at one or more sites may also be compared with model predictions via graphical displays. Invariably, the degree of correspondence between predictions and observations are described in subjective terms, an example being "...the predicted and observed wind velocities are in reasonable agreement" (Segal et al., 1982a, pg. 1392.)

Recently, Kessler and Douglas (1989) carried out qualitative performance evaluations with the Colorado State University Mesoscale Model (CSUMM) with data collected during the 1985 South Central Coast Cooperative Aerometric Monitoring Program (SCCCAMP). Four episode periods were examined. Each day was simulated separately, with the model initialized at 0400 PDT and integrated through 1200 PDT on the following day. Their evaluation consisted of visually comparing

modeled and observed winds at the surface (15 m, AGL) and at 300 m, AGL. Because the emphasis was placed on the CSUMM's ability to reproduce major features of the regional airflow rather than an accurate representation of grid-point wind measurements, the comparisons, and hence the overall performance evaluation, was subjective.

Qualitative evaluation methods are a necessary first step in the performance evaluation process. However, they provide insufficient information regarding the suitability of model output fields for input to photochemical models or whether the calculated wind and temperature distributions contain large errors or systematic biases resulting from an inadequate model, input data base, or both.

Quantitative Evaluations

A growing number of quantitative performance evaluations with urban and regional-scale meteorological models have been reported. Early studies of Pielke and Mahrer, (1978) and Simpson et al., (1979), for example, used the statistical procedures introduced by Keyser and Anthes (1977). Keyser and Anthes (1977) demonstrated the use of several simple statistics for the domain-averaged mixed layer potential temperature and sea-level pressure. In addition, they compared gridded ground-level fields of model predictions with observations. The observation field was developed by a simple interpolation algorithm (i.e. model) proposed by Keyser (1977). Investigators using Keyser's performance measures have typically considered domain-scale model performance, either averaged over the entire simulation period or for a particular hour of interest.

Tesche and Yocke (1976) introduced the use of wind speed and direction residual distributions in the evaluation the Complex Terrain Wind Model (CTWM) for photochemical model applications in the Los Angeles Basin. These residual distributions were developed for each simulation day, combining data from all monitoring stations and all hours. The mean and standard deviation of the predicted and observed speed and direction distributions were used to judge model performance. Tesche et al., (1986) evaluated the prognostic Land-Sea Breeze Model (LSBM) (Liu et al. 1979) on an hourly basis, presenting hourly bias and error estimates throughout the whole 48 hour modeling interval. Four grid-based

meteorological models were tested with detailed surface, aloft and tracer experiment data (Tesche et al., 1987). Estimates of model bias and error in wind speed prediction were computed for 16 tracer experiments. Moore et al., (1987) updated the LSBM and reported hourly bias, error, standard deviation, and correlation statistics for the period 0600-2000 LDT on 8 August, 1984 in the San Joaquin Valley of California.

Steyn and McKendry (1988) and Ulrickson (1988) incorporate the suggestions of Willmont (see, for example, Willmont, 1981; Willmont et al., 1985) to test more rigorously a hydrostatic primitive equation model. Using the CSUMM, Steyn and McKendry assess the temporal and spatial fidelity of predicted temperatures in the Vancouver, BC region with several statistical measures. They also compare modeled and observed time series of the mixed layer depth and various components of the surface heat budget. Steyn and McKendry also present plots of predicted and observed vertical profiles of wind components. Ulrickson (1988) evaluated the CSUMM in the South Coast Air Basin using the same statistical measures proposed by Willmont and implemented by Steyn and McKendry.

Recommended Numerical Meteorological Performance Procedures

Contemporary photochemical models require hourly-averaged three-dimensional wind and temperature distributions, two-dimensional mixing height fields, and various meteorological scalars including the atmospheric water content, radiation intensity, lapse rates above and below the inversion, and atmospheric pressure. Most important to ozone model performance are the wind and mixing height fields. The role of ambient temperature is becoming increasingly more important, not so much from the standpoint of photochemical model sensitivity as from the emerging strong dependence of emissions model sensitivity to temperature. Estimation of evaporative emissions from motor vehicles, for example, is highly temperature-dependent and considerable emphasis is now being placed in developing emissions models that realistically account for the time and space variations in ambient temperature throughout the airshed, for the specific simulation days of interest. Accordingly, the mechanistic evaluation procedures we recommend may be applied to meteorological model estimates of wind speed, wind direction, temperature, and mixing height.

We recommend eleven numerical measures for quantifying the performance of diagnostic and prognostic meteorological models. Following the nomenclature of Willmont (1984), they are presented in Table 4-1. These measures include:

- Mean prediction;
- Mean observation;
- Standard deviation of predictions;
- Standard deviation of observations;
- Least squares slope and intercept regression statistics;
- Root mean square error;
- Systematic root mean square error;
- Unsystematic root mean square error;
- Index of agreement;
- RMS error skill; and
- Variance rates skill.

These measures may be computed easily for wind speed, temperature, and mixing height. Because wind direction has a crossover point between 0 degrees and 360 degrees, standard linear statistical methods cannot be used to calculate the mean or standard deviation. Recent evaluations by the EPA (Turner, 1986) suggest that the method proposed by Yamartino (1984) performs well in estimating the wind direction standard deviation. Specifically, this quantity is calculated by:

$$\sigma_a = \arcsin(\epsilon) \left[1.0 + 0.1547 \epsilon^3 \right] \quad (4-1)$$

where:

$$\epsilon = \left\{ 1.0 - \left[\left(\overline{\sin a} \right)^2 + \left(\overline{\cos a} \right)^2 \right] \right\}^{\frac{1}{2}} \quad (4-2)$$

Here, a is the hourly wind direction value. Although Yamartino's method was developed for instantaneous wind direction values of a , it is expected to apply to hourly values as well.

TABLE 4-1. RECOMMENDED NUMERICAL MEASURES FOR EVALUATION DIAGNOSTIC AND PROGNOSTIC METEOROLOGICAL MODELS

Numerical Measure	Mathematic Definition	Remarks
Mean Prediction	$\bar{\phi}_p = \frac{1}{N} \sum_{i=1}^N \phi_{pi}$	N is the product of the number of simulation hours and the number of ground-level monitoring locations providing hourly observational data.
Mean Observation	$\bar{\phi}_o = \frac{1}{N} \sum_{i=1}^N \phi_{oi}$	
Standard Deviation of Predictions	$\sigma_p = \left[\frac{1}{N} \sum_{i=1}^N \phi_{pi} - \bar{\phi}_p ^2 \right]^{\frac{1}{2}}$	
Standard Deviation of Observations	$\sigma_o = \left[\frac{1}{N} \sum_{i=1}^N \phi_{oi} - \bar{\phi}_o ^2 \right]^{\frac{1}{2}}$	
Least Squares Regression Statistics	$\hat{\phi}_{pi} = a + b \phi_{oi}$	A linear least-squares regression is performed to calculate the intercept (a) and slope (b) parameters. This regression is performed for each set of hourly data to allow several other statistics to be calculated.

(Continued)

TABLE 4-1. (Continued)

Numerical Measure	Mathematic Definition	Remarks
Root Mean Square Error (RMSE)	$RMSE = \left[\frac{1}{N} \sum_{i=1}^N \phi_{oi} - \phi_{po} ^2 \right]^{\frac{1}{2}}$	
Systematic Root Mean Square Error (RMSE _s)	$RMSE_s = \left[\frac{1}{N} \sum_{i=1}^N \hat{\phi}_{oi} - \phi_{oi} ^2 \right]^{\frac{1}{2}}$	
Unsystematic Root Mean Square Error (RMSE _u)	$RMSE_u = \left[\frac{1}{N} \sum_{i=1}^N \hat{\phi}_{oi} - \phi_{oi} ^2 \right]^{\frac{1}{2}}$	
Index of Agreement (I)	$I = 1 - \left[\frac{N (RMSE)^2}{\sum_{i=1}^N (P'_{oi} + O'_{oi})^2} \right]$	<p>$P'_{oi} = \phi_{oi} - \bar{\phi}_o$, and $\phi'_{oi} = \phi_{oi} - \bar{\phi}_o$ Willmont's (1981) index of agreement has a theoretical range of 0.0 to 1.0, the latter score suggesting perfect agreement.</p>
Skill-E (SE)	$SE = \frac{RMSE_u}{\sigma_o}$	
Skill-Var (SV)	$SV = \frac{\sigma_p}{\sigma_o}$	

A "good" meteorological model will provide low values of the RMSE, explaining most of the variation in the observations. The systematic error, RMSE_s, should approach zero and the unsystematic error RMSE_u, should approach RMSE since:

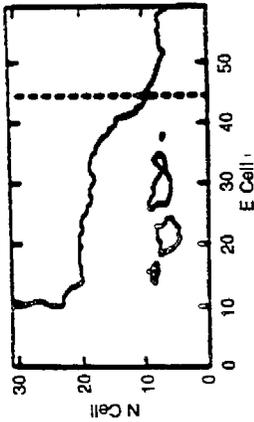
$$RMSE^2 = RMSE_s^2 + RMSE_u^2 \quad (4-3)$$

It is important that RMSE, RMSE_s, and RMSE_u are all calculated. If only RMSE is estimated (and it appears acceptable) it could consist largely of the systematic component. This bias might be removed, thereby reducing the bias introduced into the photochemical calculation. On the other hand, if the RMSE consists largely of the unsystematic component (RMSE_u), this serves as an indication to the modeler that further error reduction may require model refinement and/or data acquisition. It also provides specific error bars to be used on the meteorological inputs in subsequent photochemical model sensitivity analyses.

We recommend that all eleven measures are calculated and reported as part of meteorological model mechanistic evaluation. Currently, there is little guidance available for judging acceptable ranges of each of these measures. The recent work by Ulrickson (1988) and Steyn and McKendry (1988) provides some initial perspective, however. Ideally, a basis of experience with these performance measures will be developed over the next few years. This will lead, perhaps, to a refinement in the list of numerical measures, and to an improved perspective on acceptable ranges for each measure.

Several graphical procedures are suggested for analyzing meteorological model performance. Each of these have been usefully applied in the past. Figure 4-1 is a "slice plot" which exemplifies the resulting three-dimensional temperature fields that were input to the Land Sea Breeze Model for the 25 September, 1980 simulation. Included in the figure is the terrain elevation along a north-south slice, the vertical temperature field, and the mixing height at the hour given. Alternatively, one could plot wind vectors on the vertical slice instead of the temperatures.

Top view of the region:



----- = the slice
 Slice on cell #: +45
 Slice beg & end on cells #: 1, 31
 Date: Sept 25, 1980
 Time: 16.00

Side view of the slice:

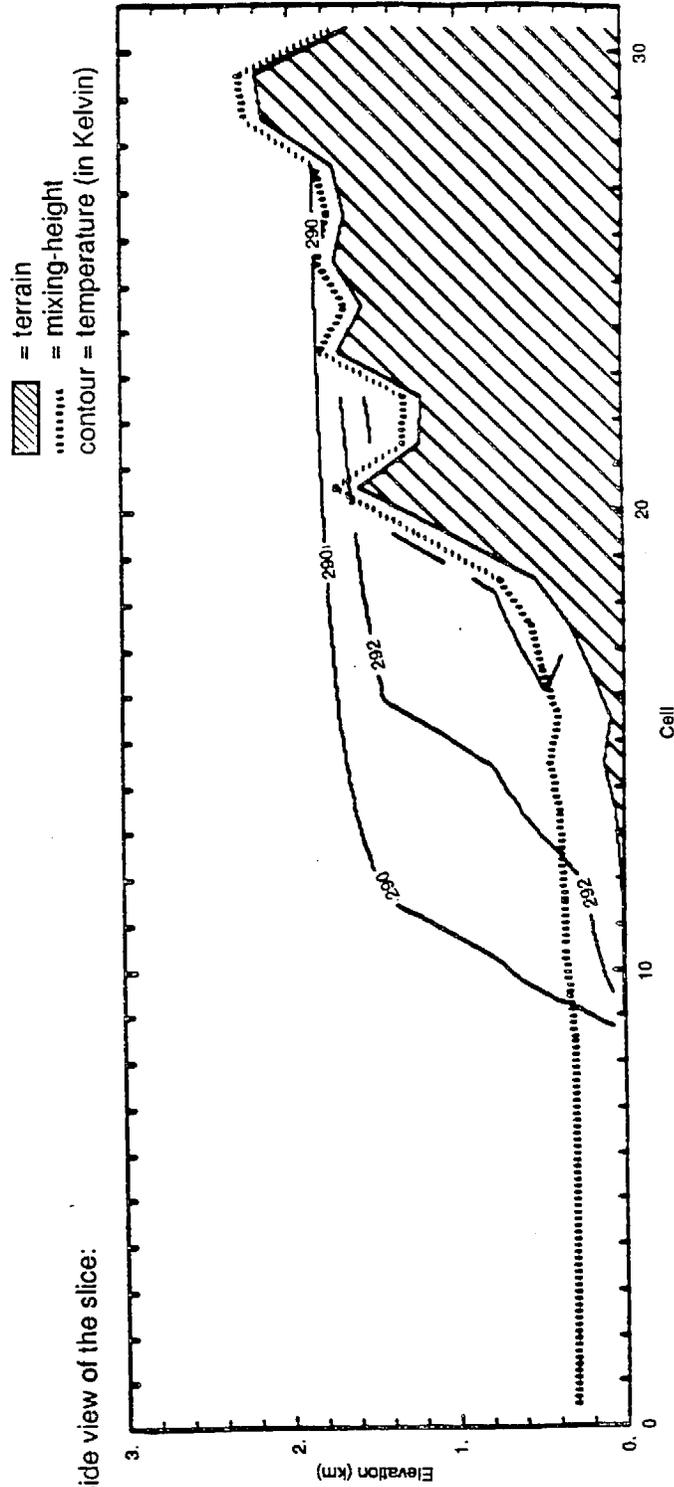


Figure 4-1. North-South Slice of Modelling Region Depicting Terrain, Temperature Structure, and Mixing Height. 1600 PST, 25 September 1980.

Figure 4-2 presents surface windfields for 0600-0700 LST on 11 September, 1980 in the SCCAB. The bold arrows correspond to surface winds observed at the monitoring stations. Graphical comparisons between predicted and observed surface wind vectors at each monitoring station for every simulation hour are often useful in diagnostic analyses. Figure 4-3 is an example of such a comparison between simulated (solid arrows) and observed (dashed arrows) wind velocities at the Ventura monitoring station on 25 September, 1980.

The results of a residual analysis of hourly predicted and observed wind speeds is shown in Figure 4-4 for three simulation days in the SCCAB. Similar plots for the wind direction residuals (not shown) are recommended. On each plot, the mean and standard deviation of the predicted and observed residual distributions should be given to aid analysis.

Figure 4-5 presents the mean wind speed, direction, wind speed standard deviation, the three root mean square errors, and the index of agreement as a function of simulation time for Ulrickson's (1988) application to 8 August, 1984 in the South Coast Air Basin. Similar plots for the standard deviation of wind direction, and the two skill metrics, SE-VAR and S-VAR, should also be plotted as a function of time. In these times series plots, the model "spin-up" interval, normally 4-6 hours at the start of the simulation, should be identified clearly. Overall performance statistics should not include results from this dynamic initialization period.

4.2.3 Chemistry Model Testing

The chemistry module is one of the few components of photochemical models that can be independently evaluated. Since the early days of photochemical modeling, scientifically sound chemical mechanisms have undergone testing against environmental chamber data before being used in atmospheric modeling (Hecht et al. 1974; Demerjian et al. 1974, 1979; Dodge 1977; Carter et al. 1979, 1986, 1988, Whitten et al. 1977; Jeffries et al. 1981; Killus and Whitten 1983; Atkinson et al. 1982; Leone and Seinfeld 1984; Lurmann et al. 1986, 1987; Gery et al. 1988; Carter and Lurmann 1989). A large body of environmental chamber data have been collected for testing chemical mechanisms. Consensus exists within

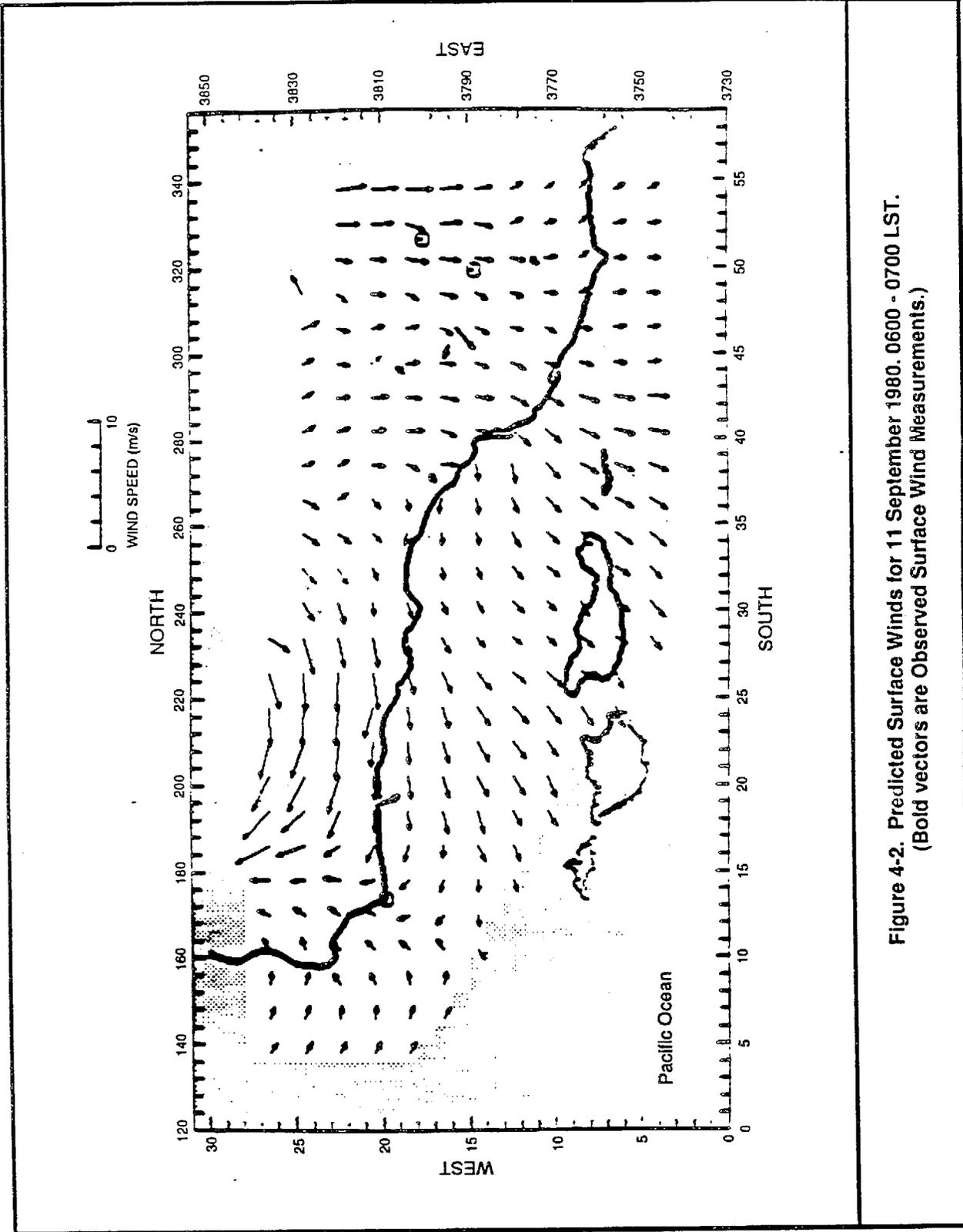
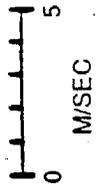


Figure 4-2. Predicted Surface Winds for 11 September 1980, 0600 - 0700 LST.
(Bold vectors are Observed Surface Wind Measurements.)

25 SEPT 1980 - SURFACE WINDS



STATION: VENTURA
 PRED ———
 OBSER - - - - -

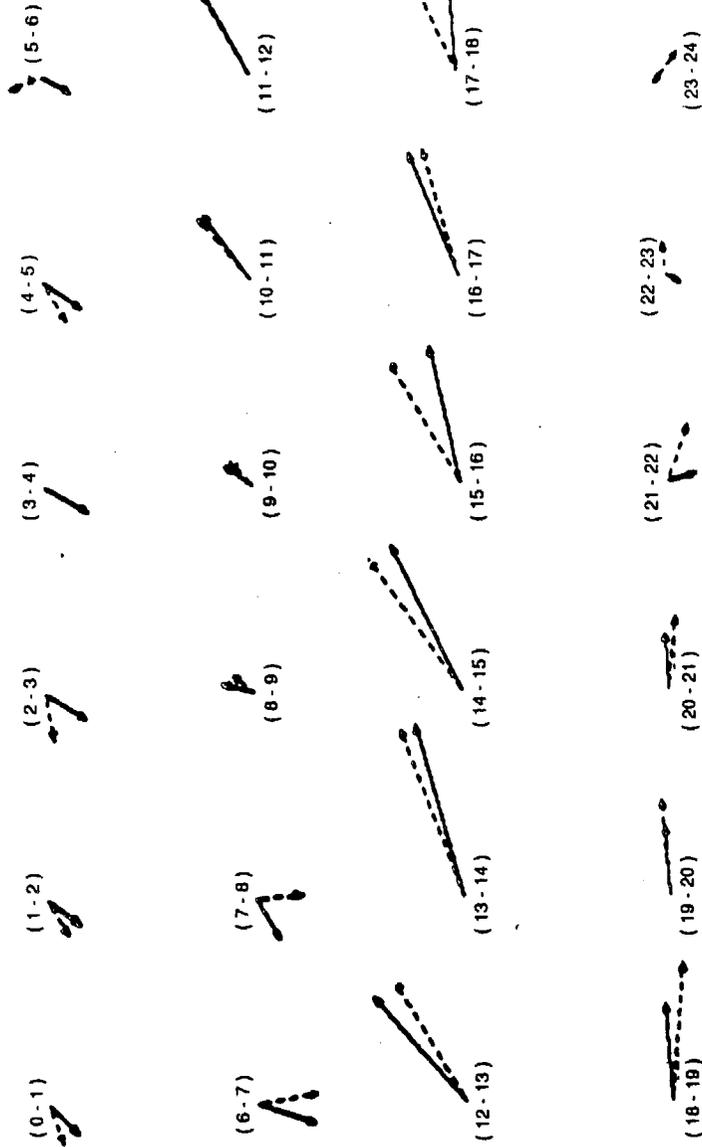


Figure 4-3. Comparison of Estimated and Observed Hourly Wind Values at the Ventura Station in the SCCAB - 25 September 1980.

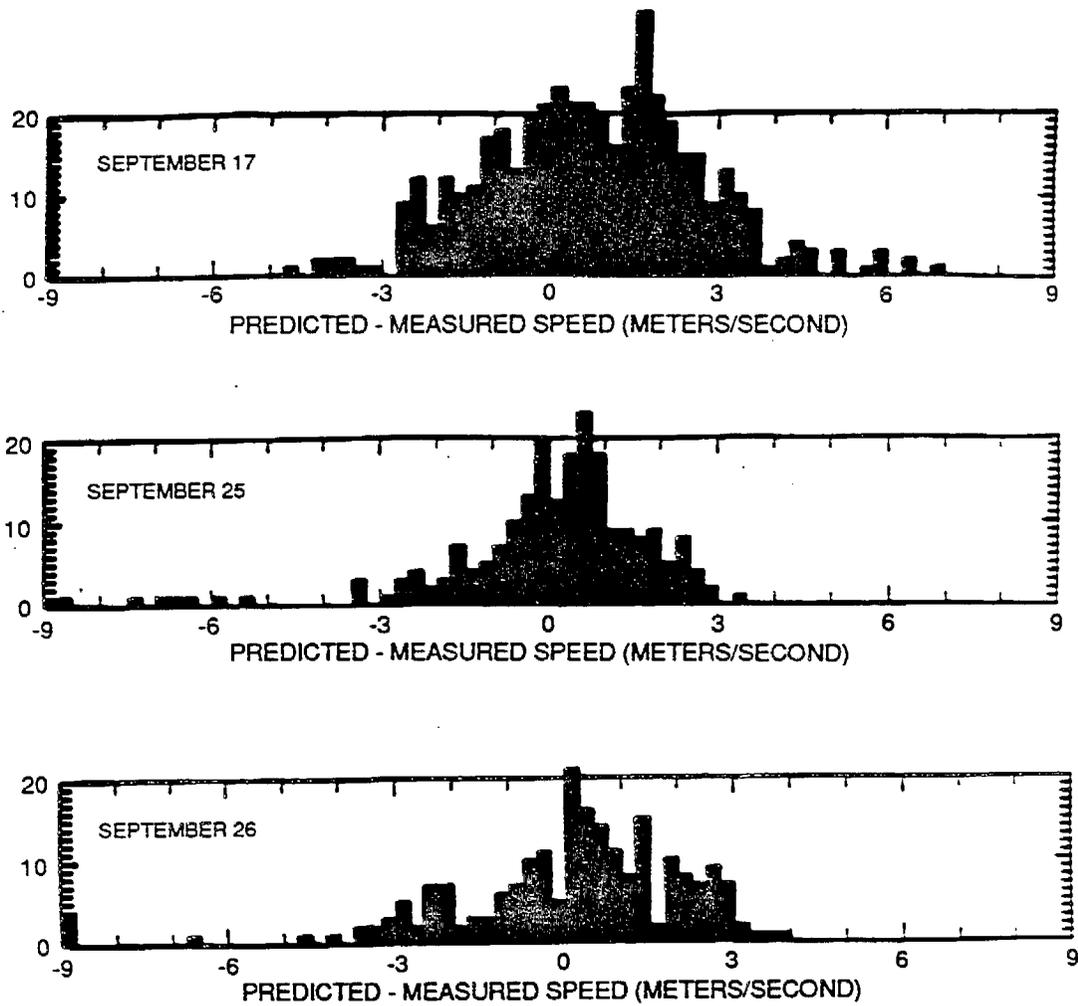
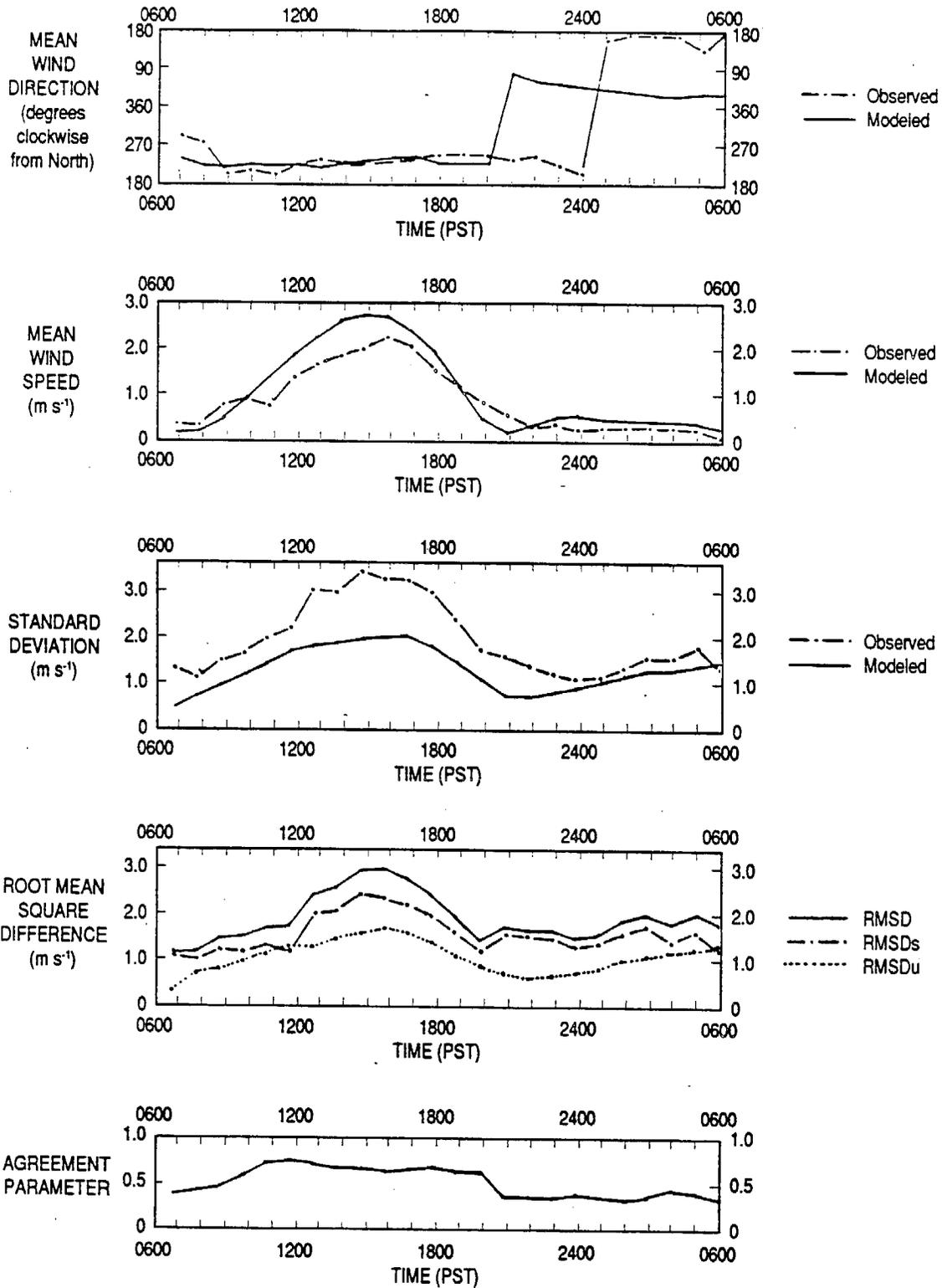


Figure 4-4. Residual Distributions of Predicted and Observed Wind Speeds.



Source: Ulrickson, 1988.

Figure 4-5. Statistical Comparison of Observed and Modeled Surface Winds with Time on 8 August 1984.

the scientific community that multi-species testing using these data is the best available method for mechanistic evaluation of the chemistry for urban photochemical modeling (Atkinson et al. 1987). In fact, because of the recognized uncertainties in chemical mechanisms, testing using chamber data should be a necessary requirement for using a mechanism. However, given the limitations of the chamber data, such testing can at best demonstrate an absence of bias and general consistency between predictions and observations. The data are insufficient to verify a chemistry module.

The limitations of the environmental chamber data should be recognized. First, the data base provides concentration-time profiles for NO, NO₂, ozone, PAN, selected hydrocarbon precursors, and some carbonyls. Comparisons of calculated and experimental ozone, NO, NO₂, PAN, VOCs, and carbonyls provide a basis for operational evaluation and limited diagnostic evaluation of the chemistry models. However, measurements are not available for many of the important radicals, intermediate species, and product species that could be used in determining whether or not a chemistry model is mechanistically correct. For example, measurement of OH, HO₂, HONO, HNO₃, H₂O₂, and organic radicals and products could be very helpful for mechanism testing. Second, the initial precursor concentrations used in environmental chambers are typically at or above those commonly observed in urban areas (e.g., [NO_x] > 100 ppb and [VOC] > 500 ppbC). Thus, data are not available to test the chemistry's performance in relatively clean environments. Third, chambers have artificial sources of reactivity which are not well characterized (Carter et al. 1982). The artificial source of radicals does not significantly influence the testing of organic species that are moderately or highly reactive. However, variability in the chamber radical source strength interferes with testing the chemistry of slowly reacting compounds, because the chamber radical source strength may be comparable to the gas-phase radical formation rate in such systems. This is of concern because about half of the VOCs emitted into urban air are slowly reacting alkanes whose chemistry can not be unambiguously evaluated using chamber data. Lastly, although some evaluations have used over 550 experiments (Lurmann et al. 1987; Carter and Lurmann 1989), statistically robust evaluations are only possible for a limited number of individual organic compounds, such as formaldehyde, ethene, propene, toluene, and xylene, and organic mixtures.

Procedures have been developed to characterize conditions in environmental chambers, such as light intensities and spectral distributions, the rates of species off-gassing and deposition on chamber walls, and the rates of heterogeneous reactions occurring on chamber walls (Carter et al. 1986; Gery et al. 1988; Jeffries et al. 1988). Significant differences exist in the chamber characterization procedures used by different research groups, and these differences may dramatically influence the results of the chemistry model evaluations (Atkinson et al. 1987; Jeffries et al. 1988).

The recommended approach for mechanism testing follows the hierarchy of species shown in Figure 4-6 (Whitten 1983; Atkinson et al. 1987). The diagram shows the hierarchical nature and interdependence of the major components of atmospheric photochemical mechanisms. Mechanism testing is initiated at the lowest level in the hierarchy (ie., NO_x-air experiments to test the inorganic reactions and the chamber characterization procedures) and proceeds to the highest level (>C₂ hydrocarbons) by step-wise addition of species with increasingly complex chemistry. The step-wise feature of this approach is designed to facilitate identification of the uncertain components of chemistry models, and minimize both the occurrence of compensating errors in the mechanism and fortuitous agreement between models and chamber data.

The rationale for testing according to the hierarchy is that if a substantial disagreement between a model and data occurs at a high level in the hierarchy and the model has shown to perform well for experiments involving species lower in the hierarchy, then the source of the disagreement is most likely in the new higher level chemistry rather in the lower level reactions. Following tests of single hydrocarbon-NO_x systems, the chemistry models are tested against complex mixtures of organics in the presence of NO_x. The complex mixtures are surrogates for the organics present in ambient air. Testing using complex mixtures allows for investigation of synergistic and compensating effects.

An important component of diagnostic evaluation is searching for relationships between model biases and input parameters. Several procedures have been used to investigate the presence of these relationships in chemistry

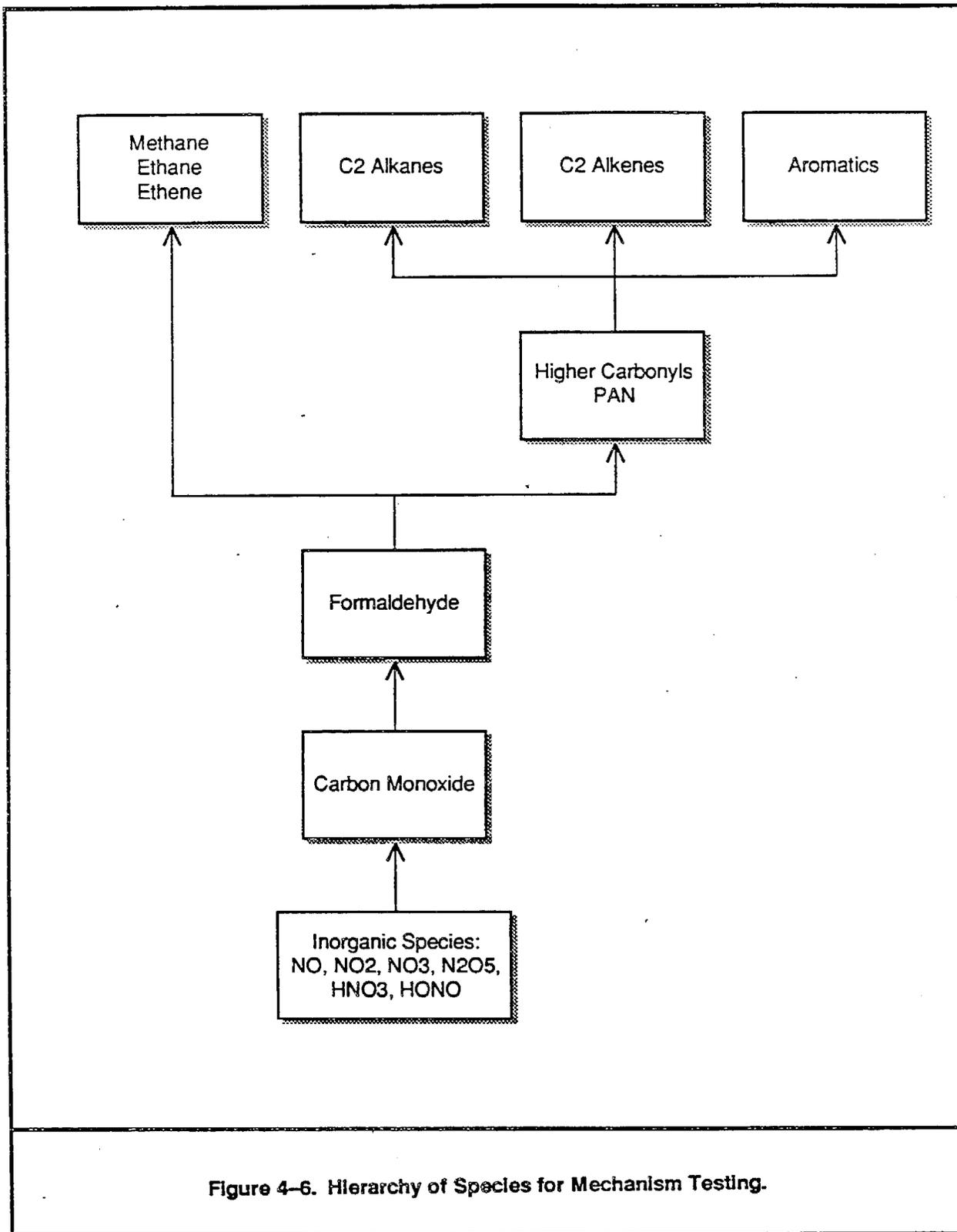


Figure 4-6. Hierarchy of Species for Mechanism Testing.

models. Perhaps the oldest method is to examine the errors in ozone as a function of the different types of organic precursors (which is now routine). Plots of the errors in ozone (or other product species) as a function of the initial NO_x and VOC concentrations, and VOC/NO_x are useful for identification of systematic biases. Figure 4-7 shows a typical plot of normalized error in ozone versus the initial VOC/NO_x in complex mixture runs. It shows under- and overprediction of ozone by comparable amounts and with comparable frequency for initial VOC/NO_x of 1 to 20. This demonstrates an absence of bias within this range. However, at the lowest and highest VOC/NO_x , this mechanism has a positive bias. If a mechanism has negative bias at low VOC/NO_x and positive bias at high VOC/NO_x (or vice-versa), it is indicative of mechanistic errors in the model that must be corrected before use in atmospheric modeling and, especially, in control scenario evaluation.

Research needed to improve mechanistic evaluation of chemistry models include: (1) acquiring chamber data from existing facilities and with existing methods for organic species for which there are little data; (2) developing and applying methods to collect data for species not presently measured in existing facilities; and (3) development of cleaner environmental chambers. The data collection and methods development efforts should focus on the following major areas of uncertainty in current photochemical mechanism: the identification and subsequent reactions of aromatic ring-fragmentation products, the radical yields in ozone-olefin reactions, the oxidation mechanisms of $>\text{C}_5$ alkanes, and photolytic data for carbonyls.

4.2.4 Deposition Model Testing

Dry deposition at the surface of an air basin is a major sink for gaseous air pollutants. Model calculations in the South Coast Air Basin show, for example, that approximately 35 percent of the NO_x emitted into an air parcel is removed by dry deposition within the first day of transport (McRae and Russell, 1984). Clearly, if the dry deposition calculations incorporated into photochemical air quality models are in error, then predictions of ozone and ozone precursor concentrations remaining in the atmosphere also will be in error.

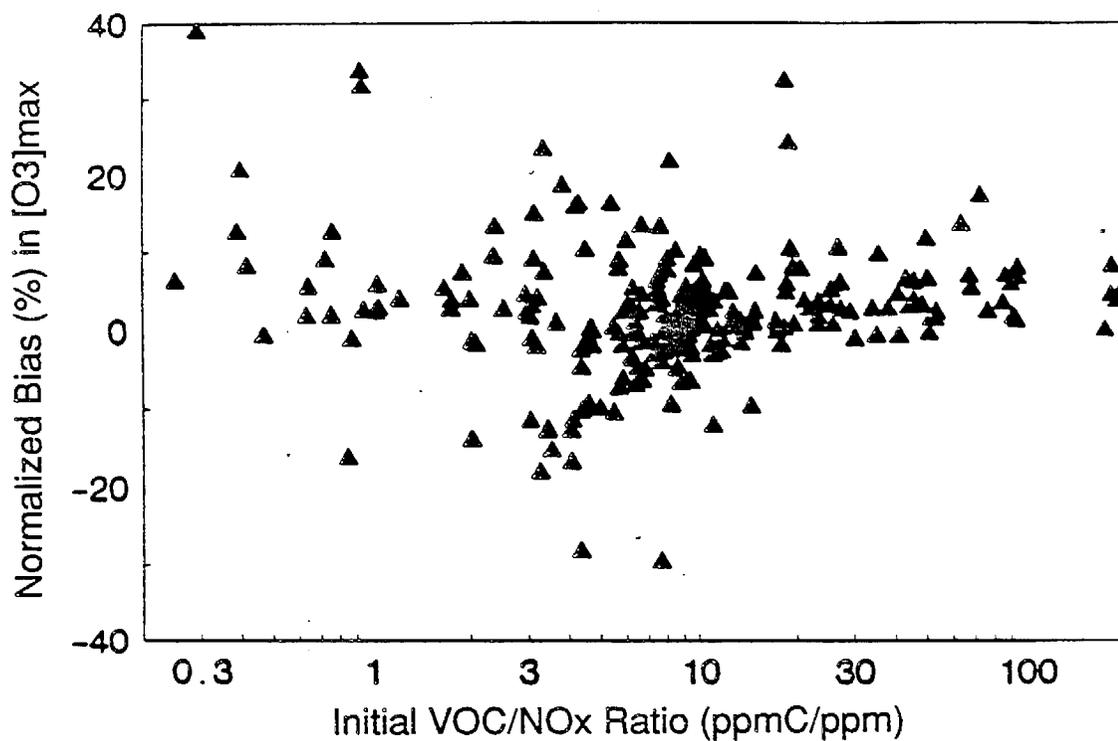


Figure 4-7. Normalized Bias (%) in Maximum Ozone in Simple and Complex Organic Mixture Runs.

The deposition flux, F , to a surface is often computed by multiplying the atmospheric concentration of the pollutant of interest by a parameter known as the deposition velocity, v_d :

$$F = v_d \cdot c(z_r) \quad (4-4)$$

where c is the pollutant concentration measured at reference height z_r . Within simplified air quality models, v_d is often specified as a constant value based on the results of field experiments in which the concentration, c , and flux, F , were measured simultaneously at the surface. In reality, the deposition velocity is not constant, but rather depends on atmospheric turbulence and surface conditions such as moisture level. Turbulent transport limits the rate at which pollutants are brought to the vicinity of the earth's surface, molecular scale diffusion of the pollutant across a stagnant atmospheric layer immediately adjacent to the surface provides another barrier to deposition, and finally, the chemical reaction characteristics of the surface and pollutant determine whether or not the gas molecules become attached to the surface.

Photochemical grid models include dynamic transport calculations and chemical reaction resistance calculations in estimating deposition rates as a function of time and location. The correctness of these calculation steps should be checked individually to the extent possible. Then, the deposition module, in isolation from the full photochemical model, should undergo mechanistic evaluation using special studies in which the atmospheric concentration, meteorological and surface characteristics inputs to the module and the flux to the surface were measured simultaneously. These tests will determine whether or not correct concentration values can be mapped into correct deposition fluxes. Finally, the deposition module as it is integrated into the photochemical model should be evaluated to determine the accuracy with which the pollutant flux can be calculated directly from emissions data and meteorological data.

Before mechanistic evaluation of a candidate deposition module, preliminary assessment first should be undertaken to ascertain whether or not

the theoretical basis of the calculation is reasonable. Key questions to be answered include:

- Is the diffusion-limited flux to the ground computed in a manner consistent with atmospheric boundary layer theory?
- Have the surface resistance values due to chemical reaction been selected based on the most recent experiments reported in the scientific literature?
- Does the model adjust for the effect of computational cell height on average pollutant concentrations such that pollutant fluxes are calculated as if the concentrations were measured at a constant reference elevation?
- Is vertical resolution sufficiently fine (or are there compensating treatments in the model) to assure that deposition is adequately simulated?
- Is vertical resolution sufficiently fine (or are there compensating treatments in the model) to assure that deposition is adequately described?

The first step in this process is to test the accuracy of the transport-limited deposition rate calculation. Experimental data should be sought on the atmospheric concentration and deposition of a pollutant that has essentially zero chemical reaction resistance to deposition. Nitric acid vapor is one gaseous contaminant that is thought to be deposited at a diffusion-limited rate (Huebert and Robert, 1985). Comparison of computed and observed HNO_3 deposition fluxes in the presence of accurate meteorological inputs to the model would constitute an initial test of the deposition module. If a sufficiently large set of such data can be acquired, the statistical measures of the bias and error for the deposition calculation can be estimated using the procedures developed in Section 2.0. Of course, the deposition rates of other species, many occurring

less rapidly, should also be tested before the deposition module can be successfully verified.

The chemical reaction resistance algorithm in most deposition models is determined empirically based on field data. Therefore, a theoretical test of this portion of the deposition calculation is not practical. A literature survey of the variability of the experimental data on the chemical reaction resistance for specific pollutant-surface pairs can be conducted. Then, sensitivity analysis procedures can be used to assess the uncertainty in the deposition module results due to the underlying uncertainty in the chemical reaction resistance values.

Emphasis over the last decade on acid deposition phenomena has led to a number of field studies in which careful measurements have been made of pollutant concentrations, meteorological conditions, and pollutant fluxes to various surfaces. Most of these studies were not conducted in geographical areas where high-resolution emission inventories are available so it is usually not possible to test photochemical models against these data sets. However, if the deposition module is isolated from the model, then the ability of the entire deposition module to track the relationship between atmospheric concentration and surface fluxes can be assessed.

The most convincing test of the deposition calculations produced by a photochemical model is analogous to the traditional testing of the concentration predictions of such a model. Surface flux predictions are made in space and in time and compared to corresponding field observations on the actual flux to the airshed. While data sufficient to support such a comparison are rare, the data are not totally absent. For example, Riggan et al (1985) have reported the nitrogen fluxes to watersheds in the San Gabriel Mountains within the grid system used for photochemical modeling in the South Coast Air Basin. Within the urbanized portion of the basin, Pierson et al. (1988) have measured the dry flux of HNO_3 to surface outdoors at Claremont, CA. Photochemical grid model predictions of the surface flux of the pollutants measured at those specific times and locations could be developed and compared against the available field observations. The existence of a few field studies of this type suggests that

research could be directed specifically toward acquiring a model evaluation data set for testing dry flux calculations over a wider geographic area and over a wider range of pollutant species.

In the absence of experimental data on the actual deposition flux, a few less exacting tests can be performed to obtain some insight into deposition model performance. Graphical tests can be considered. The spatial distribution of predicted pollutant fluxes can be contoured and examined to assure that the spatial distributions of the fluxes are in reasonable relation to the atmospheric concentrations. The ratio of the computed deposition fluxes to the computed atmospheric concentrations can be calculated and displayed spatially, resulting in a map of the effective deposition velocity for each pollutant. The deposition velocity map can be checked to ensure that the values do not exceed the fluid mechanical upper limit to the deposition velocity. Deposition velocities for pollutants with a high surface chemical reaction resistance should be lower than for pollutants that deposit at the diffusion-limited rate, and this too can be checked to verify that it is reflected in the deposition velocity maps created by the model.

Hubbe (1989) discusses several other facets of deposition flux estimation that are worth considering in developing a research plan for rigorous evaluation of deposition models. He explores ten major assumptions inherent in deposition flux modeling and discusses the results of major field measurement programs aimed at supplying data for model development and evaluation.

4.3 Non-Traditional Approaches to the Incommensurability Problem

Surface measurements of pollutant concentrations are made at instrumented monitoring stations. These measurements represent concentrations in the immediate vicinity (of the order of tens of meters to one hundred meters) of the station. In areas where concentrations vary with distance ("gradient areas"), or in areas of moderate to high variability in concentrations, measurements characterize only local conditions. In contrast, photochemical grid models estimate volume average concentrations in each grid element where the characteristic horizontal dimension of a grid cell is two to five kilometers

for urban scale applications, five to twenty kilometers for regional scale applications (such as the San Joaquin Valley), and twenty to one hundred kilometers or more for subcontinental scale models such as the acidic depositions models, RADM and ADOM. Because of the vastly different spatial and temporal scales of measurements and modeling estimates (or predictions), measured and predicted concentrations are not directly comparable. This lack of comparability, which can add an important element of uncertainty to performance evaluation exercises, is termed incommensurability.

The causes of incommensurability stem from the limitations of both measurement and modeling. Measurements made at fixed locations cannot adequately reflect variability at the spatial scale of a grid cell. This is frequently termed the representativeness problem. Conversely, grid-based models do not incorporate descriptions of processes occurring at less than grid scale. Unless either measurement or modeling practice is altered so as to bring the differing spatial scales more into alignment, the problem of incommensurability will persist.

4.3.1 Sub-grid Scale Modeling

One option for enhancing commensurability is to include in grid-based models those dynamic processes, such as chemical reaction, mixing, dispersion and deposition, that occur at spatial scales less than the size of the modeling grid. While it is possible, in principle, to develop and include descriptions of these processes, it is exceedingly difficult to do so in practice. The theoretical representations are complex and only approximate. Data needed to test the representations are difficult and expensive to acquire. Moreover, based on experience to date, there is little evidence to suggest that the performance of models would improve significantly if approximate representations of subgrid scale processes were included in models. Nevertheless, modeling of subgrid scale variability in the vicinity of monitoring stations, with the objective of achieving comparability between prediction and observation, may prove useful. However, innovative ideas and considerable time, effort, and supporting data will be required if this approach is to succeed.

4.3.2 Finer Grid Resolution

Another option for improving commensurability is to substantially reduce the dimensions of a grid cell. Incorporating increased grid resolution, however, leads to increased computing requirements (increasing at least with the square of the ratio of the horizontal linear dimension of the original to the modified grid element) and data requirements, and the need for finer scale description of dynamic processes. Moreover, for the reasons discussed by Lamb (1973), there is a limit below which reducing the grid cell size violates the assumptions inherent in the turbulence closure hypothesis of contemporary photochemical grid models. While the merits of pursuing this option might be investigated, previous assessments suggest that the costs far outweigh the benefits.

4.3.3 Remote Measurements

Potentially attractive approaches for enhancing commensurability involve changes in measurement technology and strategy. Options include remote measurement and multiple measurements within the area of a grid cell. The purpose of employing remote measurement techniques is to observe volume-averaged concentrations that might be compared directly with model predictions. EPA has developed a prototype airborne lidar system for quantitatively determining volume-averaged ozone concentrations. However, there is only one such system extant, and it is still in the "testing" stage of development. A few lidar systems are available for observing SO₂ concentrations.

Availability and use of operational, commercial remote measurement systems would undoubtedly advance model performance evaluation. Of all approaches to resolving the commensurability problem, the widespread use of remote monitoring is the most promising and attractive. However, if this capability is to materialize, EPA, ARB, and other agencies will need to commit to its successful development and implementation.

4.3.4 Multiple Measurements Within a Cell

Monitoring at multiple sites within a grid cell provides an alternate means for estimating grid-averaged concentrations. For example, several years ago the Department of Energy (DOE) deployed fifty monitoring sites in the vicinity of Fort Wayne, Indiana (Dana et al., 1984) to determine the subgrid scale variability in precipitation, deposition and wet chemistry as a part of the OSCAR field study. Unfortunately, the costs of today's preferred equipment for monitoring gaseous pollutant concentrations preclude any but the most limited research use of this option. The development of very low cost measurement techniques would, of course, completely change the cost-benefit balance.

Finally, it is conceivable that, for well defined situations such as monitoring sites located along urban arterials and freeways, correlations between local point measurements and grid volume averages might be developed through special studies. However, since short term variations in local meteorological conditions influence both local and spatially averaged concentrations, it may be difficult to develop reliable correlations even under the best of circumstances.

In summary, no approach to resolving the commensurability issue is now available. The development of economical remote measurement systems and very low cost surface monitoring devices provide the most attractive routes to resolution.

4.4 "Stressing" a Model

Oftentimes, testing of grid-based photochemical models is confined to the study of one or two episodic periods of high ozone concentrations. In California, high ozone concentrations are observed during two- to four-day periods (episodes) of high pressure, high temperature and light winds. The characteristics of adverse meteorological conditions vary relatively little for a given region in the state; unfortunately, testing model performance for only these conditions does not provide a sound basis for ensuring that a model will perform adequately over a sufficiently broad range of circumstances. Of

particular interest are conditions of reduced emissions, under which acceptable air quality is observed even though the meteorology is unfavorable.

During initial evaluation of model performance, the modeler has the opportunity to carry out diagnostic analysis of results; to seek explanations for inadequate performance; to make adjustments to the model or, more frequently, to the meteorological, air quality, or emissions inputs to the model; and to retest. Because of paucity of data or uncertainty in estimates of model inputs, it is often difficult to identify unambiguously a cause of failure; several flaws may exist and yet may not be obviously present or, if identifiable, easily separable. Moreover, adjustment of inputs for one variable may improve model performance but may not correct the inherent problem; the adjustment may simply hide the flaw, albeit inadvertently.

For example, there is currently debate over the appropriateness of emissions factors used to develop estimates of on-road motor vehicle emissions in California. The recent tunnel study in Los Angeles (Ingalls et al., 1989) reported VOC emissions factors from in-use vehicles that exceed those used by the EMFAC7D model by as much as a factor of 6.9. Whether, and to what extent the tunnel study and other recent field work in Texas, Colorado, and the SOCAB point to a systematic underestimation in motor vehicle emissions estimates is presently unknown and is deserving of immediate, focused investigation. But, it is known that photochemical grid models have a systematic tendency to underestimate hourly-average ozone concentrations (Tesche, 1988). If it is determined that the motor vehicle emission model is flawed, it opens to question the reasonableness of other modeling procedures that may have, inadvertently, compensated for a negative ozone prediction bias. Specifically, have mixing height or wind speed estimation methodologies erred in favor of producing higher ozone model prediction as the direct (and unknown) result of a systematic underestimation in emissions?

In general, in carrying out photochemical grid modeling there is significant risk of having a model appear to predict adequately, but to do so for the "wrong" reasons. Under test conditions, one or more hidden inadequacies may not reveal themselves. However, performance failures may become manifest

under the reduced emissions conditions associated with implementation of a control strategy. If this situation exists, then the model will predict incorrectly, either under- or overestimating the concentration reductions associated with the strategy under study.

"Stressing a model" is designing and carrying out performance tests that cause a model to reveal its flaws and weaknesses if it is indeed inadequate. Stressful testing is intended to reduce (or avoid) the risks associated with "weak" or otherwise inadequate tests, wherein a model is not challenged sufficiently to reveal its flaws and weaknesses or appears to be performing acceptably despite significant inadequacies in formulation or inputs.

The notion that a model should be forced to reveal its flaws and weaknesses should not be construed as being negative or destructive of the model development process. Rather, it reflects the fact that a model cannot be proven accurate; it is not possible to examine all the situations of interest or to collect all the data needed for any one situation. Model acceptability is demonstrated through a lack of inadequate performance over a range of conditions and tests. Confidence in the model builds as the number and variety of "successful" tests increases. The more stress that testing places on a model, the greater is the value of the tests, and a fewer number of tests is needed to establish model acceptance.

Tests intended to stress a photochemical grid model must clearly transcend operational performance evaluation exercises, typically, the comparison of predictions and observations for ozone (and, with less stringency, for NO₂) for one or two episodes of two days' duration. No prescription of "stress testing" now exists; however, one can conceive a number of requirements which, if adopted in whole or in part, would certainly reduce the risk of inappropriately accepting a flawed model for use in control strategy assessment. For example, one might require, as part of a performance evaluation protocol, to:

- Give equal, or nearly equal, weight to accurate estimation of NO₂ and volatile organic compounds as that given to ozone.

- Attempt, insofar as possible, to assure commensurateness in space and time of predictions and observations; that is, attempt to assure that observations are representative of predictions.
- Test the performance of individual model components, such as the chemical mechanism or deposition algorithm, prefatory to examining overall model performance, whenever possible.
- Test the performance of "preprocessor models", notably the meteorological and emissions models, whenever possible.
- Examine performance of the model over the range of meteorological, air quality (chemical), and emissions conditions available for study. If performance evaluation is part of a larger program involving field observations, design the program to acquire the necessary data. For example:
 - (a) While adverse meteorological events are typically studied, evaluate model performance for situations involving cleaner air;
 - (b) Carry out tests for weekend as well as weekday emissions patterns; and
 - (c) Carry out tests for adverse meteorological conditions of varying temperature maxima and diurnal temperature profiles.
- Require "hands off" testing after an adequate period of time is allowed for the "hands on" testing that is traditionally carried out during model development. Also, require testing by independent parties (who have no vested interest in the outcome) after the inception of the period of "hands off" testing.

When a flaw or weakness is revealed through stressing the model, the intent is to correct the problem and retest. Test results are interpreted and attempts are made to identify model components and/or data preprocessing that need improvement. The process of testing and interpretation of test results is cyclic and iterative; test failures are diagnosed and improvements are made to the model until the model passes the prescribed tests, interpretation of test results can no longer reveal failed model components, or the model cannot be improved in the near term (signaling long-term model development activities and eventual re-evaluation, or acquisition of data bases more suitable for use in testing).

However desirable it is to stress a model, tests should nevertheless be "reasonable." Thus, a test should not attempt to demonstrate that a model cannot do what it was not designed to do. In other words, the testing process should not include requirements that are not possible to meet, either because the necessary data cannot be acquired or the model itself does not include the necessary features. For example, since urban and regional models are not designed to simulate subgrid scale processes, tests should not be prescribed that require accurate subgrid scale predictions of spatially variable parameters.

One might extend the requirement for "reasonableness" to the testing of nonessential features of a model. On one hand, if it is a nonessential feature that appears to fail, it is still an indication that all is not right with the model. On the other hand, the desire is to develop a useful model for guiding the thinking of decision-makers in the short and intermediate terms. Thus, if the model meets all current needs, it may be judged acceptable even if one or more nonessential elements are flawed. Of course, it is unlikely that the scientific community will wish to accept a model displaying these characteristics; however, it may still serve its intended shorter term function adequately.

Finally, as suggested earlier, tests should not require data that cannot be acquired. For example, while it is highly desirable to collect free radical data in the field, it is currently not possible to do so, except by using research equipment under special circumstances. The design of tests should reflect such limitations.

Anticipating potential problems with or limitations of models is a valuable aid in the development of stressful and informative tests. A careful look at past experience in analyzing model failures, coupled with a knowledge of the treatment of phenomena that are key to addressing issues of concern, gives focus and direction to the test development process. Possible limitations of models in their simulation of key processes, as identified by the scientific/modeling community, should guide the design of tests.

4.5 Use of Simplified Models

In certain, carefully selected situations, the use of simplified models may augment the procedures used to evaluate the full photochemical grid model. A few examples are discussed below.

4.5.1 Photochemical Box Models

Photochemical box models cannot be used to predict the magnitude and location of the concentration of photochemically reactive species in a given region; at best they may be able to calculate the temporal variation of the average regional concentration for the various pollutants. However, since they lack spatial resolution, they cannot be used in situations where the meteorological or emission patterns vary significantly across the modeling region. Clearly, single-box models cannot be used to assess the effectiveness of emission control strategies in relation to spatially inhomogeneous emissions. Nevertheless, box models that incorporate a full-scale chemical mechanism can be used in a supplemental manner in the process of evaluating the performance of a grid-based photochemical model, always for applications with respect to a given region. For example, box models can be used to calculate material balances over an entire region, and for preliminary sensitivity testing, to determine the relative contributions of sources, initial conditions, and inflow to a region's air quality.

4.5.2 Photochemical Trajectory Models

Trajectory models encompass a wide variety of formulations (uniformly mixed versus vertically and/or horizontally resolved control volume or "air parcel," constant versus expanding control volume, forward versus backward trajectory calculations). In general, the formulation employed by the trajectory models to describe atmospheric pollution dynamics represents an attempt to solve the mass conservation equation in a moving coordinate system. The air parcel of interest is assumed to travel solely with the prevailing horizontal wind and to retain its vertical shape as it is advected; actual winds are approximated by a mean wind velocity that is assumed to be constant with height. Hence, vertical wind velocities and wind shear are neglected. Trajectory models typically assume that the horizontal dimensions of the air parcel remain constant and unaffected by convergence and divergence of the windfield. Many trajectory models ignore horizontal dispersion effects but consider the dynamics of a several vertically stacked cells. Other models assume uniform mixing in the vertical direction. The entrainment of ambient air into the air parcel is also often ignored; still, even those trajectory models that include entrainment of ambient air may not account for the chemical evolution of ambient air prior to its entrainment. The practices of neglecting vertical winds, wind shear, dispersion processes, and possibly entrainment processes, and of assuming coherent air parcels, can have significant impact on the reasonableness of the predictions, especially in the case of complex windfields and in calculations for long downwind distances.

The limitations of trajectory models that are a consequence of the above practices have been detailed elsewhere (e.g., Liu and Seinfeld, 1975; Ghim and Seinfeld, 1988; Seinfeld, 1988). The usefulness trajectory models is in quickly evaluating the effect of input parameter changes on pollutant levels, i.e., for efficient, low-cost, sensitivity analyses. Although trajectory models can incorporate nearly all the chemical detail of a grid-based model, they do suffer from a number of serious and well-known limitations that restrict the applicability of the model concept. In particular, trajectory models are inappropriate for situations involving complex terrain, overnight carryover of pollutants, flows where the parcel assumption is violated (e.g., wind speed or direction shear),

and regions where strong concentration gradients occur. Yet, in some simple cases, one may estimate the sensitivity of ozone predictions to variations in initial conditions, emissions, and upper level boundary conditions with this class of photochemical model (see, for example, Lurmann et al., 1986; Derwent and Hov, 1988).

A low-cost screening (or preliminary) analysis can often be performed using trajectory models. Such an analysis should employ EPA Guideline models (i.e., RPM and PLMSTAR) that incorporate up-to-date versions of the chemical mechanisms and of the descriptions of removal processes. The quality of the input data used for applying these models should be comparable to that expected to be used with the grid-based model. The objectives of this type of analysis should be limited and well defined: it should focus on identifying questions to be answered through a full scale performance evaluation procedure.

Sensitivity analyses using a trajectory model may focus specifically on the chemical transformation and the removal (deposition) components of the photochemical model. The trajectory model should use the same chemistry and removal components (modules) as the grid based model and the same constitutive parameters (reaction rates, deposition velocities, etc.). Not only the conceptual framework for these components but also the algorithmic or numerical implementation should be identical or equivalent for the two models. The complete input parameter set of the trajectory model (involving windfield components, dispersion parameters, etc.) should be a subset of the corresponding input set for the grid model. The same vertical resolution and the same cell size should be used for both models.

A suggested list of sensitivity tests that can be performed using a trajectory model follows: it includes typical factors for calculating high and low values for the sensitivity variables from the baseline case values.

- Low (x .33) and high (x 1.5) ozone aloft;
- Low (x .5) and high (x 2) VOC aloft;
- Low (x .1) and high (x 10) NO_x aloft;
- Low (x .5) and high (x 2) initial VOC;

- Low (x .2) and high (x 5) initial NO_x;
- Low and high initial VOC and NO_x (together/factors as above);
- Low (-30%) and high (+30%) VOC and NO_x emissions;
- Low (x .1) and high (x 3) initial aldehydes surface and aloft;
- Low (x .2) and high (x 5) HONO formation rate; and
- Zero dry deposition.

4.5.3 Suggested Uses of Simplified Photochemical Models

Simplified simulation models such as photochemical box models and trajectory models (including EKMA) should not be in general considered as alternatives to grid-based photochemical models. The reason for this lies in the severe limitations inherent in the formulations of these simplified models (see, e.g., Seinfeld, 1988). The potential usefulness of trajectory models is their utility in (1) performing a preliminary analysis of the application(s) relevant to which the grid-based model is going to be evaluated, to identify aspects of the modeling task that need to be studied in detail and therefore are appropriate as focal points of the full scale performance evaluation process, and (2) in quickly evaluating the effect of input data and parameter changes on pollutant levels, i.e., in performing efficient, low-cost, sensitivity analyses specifically relevant to the transformation/removal components of grid-based photochemical models. Box models can be used to perform mass balances and check sensitivity to boundary and initial conditions for the entire region covered.

Simplified formulations that manage to retain certain of the salient characteristics of comprehensive photochemical models may be developed as adequate substitutes to photochemical grid models for certain limited, well-specified, sets of conditions. The methodologies underlying these potential developments can be found in classical chemical reactor theory which offers a wide range of semi-empirical approaches for non-ideally mixed continuous flow reactors. The main idea is to replace the well mixed cell with a simple network of cells (in series and/or in parallel) and appropriately assign mixing and chemical processes in these cells. Such formulations, still at the research stage, have been proposed for continental scale applications (e.g., Jacob et al., 1989). Possibly, some alternatives that represent compromises between large-scale, grid-based models

and trajectory (including EKMA) models may eventually prove useful for urban and regional scale applications and for a variety of sensitivity analyses.

5.0 SUMMARY AND RECOMMENDATIONS

5.1 Summary

Urban ozone modeling techniques have matured significantly since the original development work in the South Coast Air Basin two decades ago (Roth, 1988). Possibly the most pressing challenge today is determining whether a photochemical model performs well enough for use in regulatory decision-making. For many years, there have been no standardized procedures for conducting model performance evaluations or agreed-upon criteria for rejecting simulation results (Fox, 1981; Roth et al., 1989). The California Air Resources Board recently sponsored this study to establish the basis for consistent photochemical grid model performance evaluations in the near term and to provide a framework for performance evaluation research over the longer term. The specific study objectives include: (1) develop improved evaluation procedures for photochemical grid models, (2) develop evaluation procedures supporting regulatory decision-making concerning ozone control programs, (3) develop new procedures consistent with improved aerometric data bases such as the Southern California Air Quality Study (SCAQS), and (4) exemplify the use of the recommended procedures and performance measures.

A specific set of statistical and graphical performance evaluation procedures were developed for operational use in photochemical modeling studies in California. Included is a group of diagnostic simulations and tabular and graphical examples of ways in which the results might be analyzed portrayed.

Diagnostic model evaluation methods were also developed. While often coupled with operational performance evaluations, these strongly recommended procedures are intended to develop greater insight into the strengths and weaknesses of a particular model, data base and photochemical simulation than is afforded by the routine procedures. They seek to provide a method for stress testing models and for causing flawed models to fail in a demonstrable manner. In complex modeling situations or when the operational performance evaluation

results are suspect, these diagnostic procedures are an essential component of the overall evaluation process.

Emergent and longer term performance evaluation issues and research needs were addressed. In some cases, these needs have been identified for a some time; only recently have data bases and numerical modeling techniques become available to deal with them. In time, some of the extant research needs will be addressed successfully, leading to an improved set of model and data base testing procedures.

5.2 RECOMMENDATIONS

5.2.1 Operational Evaluation Procedures

Several numerical and graphical procedures are developed for assessing the performance of grid-based photochemical dispersion models. The methods suggested include the calculation of peak prediction accuracy indices, statistics based on concentration residuals, and time series of predicted and observed hourly concentrations. Graphical procedures are also suggested to complement the numerical measures, providing additional insight into model performance.

Numerical Performance Procedures

Ten numerical measures, involving specific comparisons between hourly predictions and observations, may be used to typify a model's overall performance in a photochemical simulation. While these measures may be applied to any primary or secondary pollutant for which adequate monitoring data are available, for the present, their use is recommended primarily for ozone and NO₂ concentrations. The recommend numerical performance measures are as follows.

Paired Peak Prediction Accuracy. The paired peak prediction accuracy, A_{ps} , examines the discrepancy between the magnitude of the peak one-hour average

concentration measurement at a monitoring station, $c_o(\hat{x}, \hat{t})$, and the concentration at the same location, \hat{x} , and at the same time, \hat{t} . The predictions and observations are thus "paired in time and space." The paired peak prediction accuracy is a very stringent model evaluation measure. The predicted concentration recommended for numerical all comparisons with observations is the four cell weighted average determined by simple bilinear interpolation among the four grid cells nearest the monitoring location.

Temporally-Paired Peak Prediction Accuracy. The temporally-paired peak prediction accuracy, A_t , examines the model's ability to reproduce the highest observed concentration in the subregion surrounding the monitoring station at the same time of occurrence of the measured maximum. Relaxation of the spatial-pairing requirement (compared with A_s) is allowed, up to a maximum subregional distance of 25 km.

Spatially-Paired Peak Prediction Accuracy. The spatially-paired peak prediction accuracy, A_s , describes the discrepancy between the magnitude of the peak one-hour average concentration measurement at a monitoring station and the highest predicted concentration at the same monitor, within three hours of the peak. When interpreted along with other measures, A_s provides some insight into the reasonableness of the simulated transport processes leading to the maximum concentration.

Unpaired Peak Prediction Accuracy. The unpaired peak prediction accuracy, A_u , describes the difference between the magnitude of the peak one-hour average observed concentration and the highest value predicted anywhere in the modeling region. This is the least stringent of the peak prediction accuracy measures.

Average Station Peak Prediction Accuracy. The average station peak prediction accuracy, \bar{A} , is the mean of the spatially-paired peak prediction accuracies averaged overall monitoring station locations. \bar{A} is calculated by

first determining the spatially-paired peak prediction accuracy at each monitoring station, $A_{s,i}$. The average station peak prediction accuracy is simply the algebraic mean of the absolute value of the $A_{s,i}$ scores. The temporal offset between predicted and observed maximum at any monitoring station should not exceed three hours. The average station peak prediction accuracy describes how well the maximum concentrations throughout the monitoring network are reproduced.

Mean Bias. The mean bias (i.e. mean bias error) is calculated both as a residual quantity and one that is normalized by the observed concentrations. The bias is determined from the average signed deviation of the concentration residuals. It indicates the degree to which predicted one-hour concentrations are over- or underestimated. Based on the ensemble of prediction-observation pairs, this measure reveals the presence of systematic deviation from observed concentrations. The non-normalized bias is calculated to aid in developing a robust data base on photochemical model performance evaluation. The mean normalized bias, generally of greater interest, is useful in identifying systematic errors in the model's temporal or spatial response. Since the bias reveals the tendency for systematic over- or underestimation, it should be zero in the ideal case. Caution must be exercised in the interpretation of bias because it is possible for large, compensating subregional biases to produce a mean zero estimate.

Variance. The variance of the distribution of residuals, S_0^2 , describes the dispersion of the residual distribution about the mean. As the second moment of the concentration residual distribution, the variance is a measure of the average "spread" of the residuals, independent of any systematic bias in the predictions. The variance provides no direct information about subregional errors or about large discrepancies occurring within portions of the diurnal cycle.

Gross Error. The gross error, reported both as normalized and non-normalized measures, describes the average absolute signed deviation of the concentration residuals. It indicates the average (signed) discrepancy between

hourly predictions and observations, and is one of the most useful measures for comparing different model simulations. The normalized gross error is a robust measure of overall model performance, representing the average error in prediction.

Graphical Performance Procedures

Certain features of a photochemical grid model simulation are best analyzed through graphical means. In addition to revealing important qualitative relationships, graphical displays can also be used to give quantitative information. Nine different graphical methods are recommended that provide the following information:

- The relationship between the various measures of peak prediction accuracy;
- The temporal correlation between predictions and observations;
- The spatial distribution of predicted concentration fields;
- The correlation between hourly pairs of predictions, observations and residuals;
- The variation in bias and error estimates as functions of time and space; and
- The degree of mismatch between model predictions and point measurements.

Accuracy Plot. Two accuracy plots are recommended; one depicts relationships between the five numerical peak prediction measures, and the other plot provides a comprehensive summary of the peak prediction accuracy at all monitoring stations.

Time Series Plots. Probably the most useful graph for depicting photochemical model results is the time series plot. Developed for each monitoring station for which observed concentrations are available, this plot presents the hourly predictions and observations throughout the simulation period. The absolute value of the concentration residual value is also presented on the same plot. One may determine the model's ability to reproduce the peak prediction, the presence or absence of significant bias and errors within the diurnal cycle, and whether the "timing" of the predicted concentration maximum agrees with the observations. By including the residual plot on the same graph, prediction biases are more apparent.

Spatial Time Series Plots. Spatial time series plots provide information about the degree to which model discrepancies result from the procedure for selecting the predicted values. Time series plots are constructed for each monitoring station by plotting the hourly observations together with three sets of model predictions:

- The four cell weighted average ("weighted") based on bilinear interpolation;
- The prediction in the grid cell containing the monitor ("cell");
and
- The prediction closest in magnitude to the observed value, where the prediction at a given hour is drawn from one of the four nearest grid cells ("best").

The spatial time series plot provide useful diagnostic information about the "steepness" of the concentration gradients in the simulated fields. Spatial time series plots are one method of revealing the commensurability between volume-averaged model predictions and point measurements.

Ground Level Isopleths. Ground-level isopleths display the spatial distribution of predicted concentration fields for any selected hour. Developed by computer-contouring the hourly, gridded model predictions, these isopleths supply direct information about the magnitude and location of pollutant concentrations and help to identify situations where subregional biases may be attributed to spatial misalignment of the predicted and observed concentration fields.

Scatter Plot of Predictions and Observations. Scatter plots depict the extent of bias and error in the ensemble of hourly prediction-observation pairs. Bias is indicated by the preponderance of data points falling above or below the perfect correlation line. The dispersion of points is a measure of error in the simulation. Scatter plots are helpful in identifying potential outlier prediction-observation pairs. These plots provide little diagnostic information about subregional performance problems, temporal or spatial misalignments, or other inadequacies in the simulation.

Scatter Plot of Residuals and Observations. The residual scatter plots reveal the distribution of hourly model discrepancies (positive and negative) as a function of concentration level. The plot does not reveal the existence or causes of subregional or timing performance problems. The smaller the scatter about the ordinate, the smaller the modeling error. Absence of bias is indicated by no systematic tendency for the data points to fall above or below the ordinate.

Bias Stratified by Concentration. The bias-concentration plot depicts the degree of systematic bias in hourly-averaged model predictions (paired in time and space) as a function of observed concentration level. The bias-concentration plot reveals the existence of under- or overestimation within any concentration interval.

Gross Error Stratified by Concentration. The gross error-concentration plot depicts the degree of error in model prediction (paired in time and space)

as a function of observed concentration level. The gross error-concentration plot, revealing the variation in model error at various intervals throughout the concentration range, must be interpreted carefully because the residual error is normalized by the observed concentration.

Bias Stratified by Time. The bias-time plot identifies specific time periods within the photochemical simulation when systematic tendencies toward under- or overestimation occurs. The bias-time plot is constructed in a manner similar to as the bias-concentration plot, except that the simulation period is discretized into a number of time intervals, usually 1-2 hours in duration.

Gross Error Stratified by Time. The gross error-time plot identifies specific time periods when gross errors in the model predictions may be a problem. The gross error-time plot is constructed in a similar manner as the error-concentration plot.

Diagnostic Simulations

Six diagnostic simulations are suggested as the minimum set of tests to accompany the numerical and graphical procedures previously outlined.

Zero Emissions. The purpose of the zero emission diagnostic simulation is to ensure that the base case simulation results are influenced appropriately by the emissions inputs. Zeroing out all emissions should lead to significantly-reduced reactive species concentrations on the second and subsequent simulation days. The zero emissions simulation is performed by exercising the base case run with all emission values reduced to zero. All other model input files remain unchanged from the base case.

Zero Initial Conditions. The zero initial conditions simulation quantifies how much of the second (or third) day predictions are the direct result of the initial conditions used to start the simulation. This simulation is performed by setting all initial concentration fields in the model to zero.

If the initial field is completely "washed out" of the model domain by the second or third day, simulation results will indicate essentially no differences between the diagnostic and base case runs on the following day(s).

Zero Boundary Conditions. The purpose of the zero boundary condition simulation is to quantify the influence of boundary conditions on second (or third) day concentrations, particularly in regions where the base case predictions are highest. This simulation helps identify situations where the base case results are "driven" by the boundary conditions. The zero boundary conditions simulation is performed by setting all inflow and outflow boundary values, including the region top, to zero. Unless there is reason to suspect that a portion of the peak concentration measurements within the region derive from transport from outside the modeling region, the model results should reveal little impact of the boundary conditions in the interior of the computational domain.

Zero Surface Deposition. The zero deposition simulation quantifies the influence of dry surface deposition removal on primary and secondary species concentrations. The zero deposition diagnostic run is exercised by setting deposition velocities for all species to zero and re-running the base case simulation. Deposition tests have not been reported in previous model evaluation studies so the analyst has little historical information at present to serve as a guide in interpreting the results of this investigation.

Increased Mixing Heights. The objective of the mixing height diagnostic simulation is to reveal the degree to which ozone concentrations are influenced by the height of the mixed layer. At a minimum, one run is suggested in which the hourly mixing height values are uniformly increased by 50 percent above the base case values. This diagnostic simulation should provide a bound on the change in ozone predictions resulting from uncertainties in this input. One might choose, instead, to reduce the hourly mixing heights by 50 percent. The resultant increase in ozone concentrations under this scenario will typically

be comparable in magnitude but of opposite sign as those for the mixing height increase case.

Reduced Wind Speeds. This diagnostic simulation entails a 50 percent reduction in the magnitude of the winds input to the photochemical model, thus providing an initial characterization of the ozone model's sensitivity to ventilation. This simulation is performed by decreasing all grid-point wind components by 50 percent and re-running the photochemical model. The magnitude of the hourly ozone concentrations, including the peak value, should increase relative to the base case although the percentage increase should be less than proportional with wind speed reduction.

Guidelines for Assessing Simulation Results

We do not endorse setting rigid criteria for model "acceptance" or "rejection" largely for the technical and policy reasons discussed by Fox (1981) and Roth et al., (1988). Instead, the following approach is suggested. From over 15 years of photochemical model development and testing, photochemical grid model simulations generally produce peak (unpaired) prediction accuracy, overall bias, and gross error statistics in the approximate ranges of 15-20 percent, 5-15 percent, and 30-35 percent, respectively. A study following an approved Ozone Modeling Protocol that falls below all of these ranges would not be rejected unless evidence from the model sensitivity simulations and the other numerical measures and diagnostic tests suggest unusual or aberrant behavior. For model simulations falling within these ranges, some additional diagnostic analyses (such as those proposed in Sections 3.0 and 4.0) may be appropriate to lend further support to the contention that the simulation is acceptable. For model results outside any one of these general ranges, it should be incumbent on the modeler to explain why the performance is poorer than that commonly achieved in similar applications. The modeler should also explain whether the causes of poorer performance will adversely affect the use of the model in control strategy evaluations. This methodology provides reviewing agencies with a general model performance target, but still guards against the inappropriate rejection of less

accurate model simulations when appropriate and explainable reasons can be provided.

5.2.2 Optional but Strongly Recommended Evaluation Procedures

Sensitivity analysis is an essential component of model performance evaluation and should be conducted as part of the overall study. Sensitivity analyses help to reveal internal inconsistencies in the model, identify the inputs that dominate the model's operation, clarify error propagation through the model, and their results help develop guidance for model refinement and data collection programs.

Sensitivity-Uncertainty Analysis

A sensitivity analysis should normally accompany a major model performance evaluation study. The key input variables and parameters need to be identified and their levels of uncertainty estimated. These uncertainty bounds are then propagated through the model, either singly or in concert, to provide estimates of the uncertainties in the predictions. The modular should discuss the results of the sensitivity runs from the viewpoint of the response of the model expected from its underlying physics and chemistry. Typical inputs and parameters that should be considered for sensitivity analysis include:

- Key emission parameters, such as VOC rates and speciation;
- Mixing height;
- Wind field; and
- Deposition velocities.

Ideally, comparisons of predictions and observations should account for uncertainties in both observational data and model predictions. One then seeks to determine if the uncertainty envelopes around the data and the predictions overlap sufficiently. The prediction envelope is generated by a

sensitivity-uncertainty analysis. We recommend that performance statistics when reported reflect this concept.

Multi-Species Comparisons. The development of evaluation procedures that test photochemical model performance for species other than ozone has been strongly recommended by the Urban Airshed Model Technical Advisory Committee (Seinfeld et al., 1987) and the SCAQS Model Working Group and Data Analysis Group (Seinfeld et al., 1988). The reasoning for these recommendations is that multi-species comparisons can provide a more robust basis for accepting or rejecting a model (or a model simulation); they significantly improve the chances that a flawed model will be identified. Adequate model performance for several reactive species increases the decision-maker's assurance that correct ozone predictions are not a result of chance or fortuitous cancellation of errors introduced by various assumptions. We recommend in addition to ozone the following species for performance evaluation if suitable data are available:

- NO, NO₂, VOC, VOC speciated;
- H₂O₂, HCHO, PAN; and
- HNO₃, particulate nitrate.

Mass Fluxes and Budgets. Only recently have attempts been made to derive mass balances and carry out flux calculations for photochemical grid model simulations. Four mass balance and flux calculations are suggested to accompany detailed performance evaluations. These include:

- Mass fluxes into and out of domain boundaries;
- Mass fluxes into and out of the mixed layer;
- Surface deposition fluxes; and,
- Reconciliation of emission, transport, transformation and removal terms in a closed budget.

Mass balances and flux calculations have been performed only to a very limited extent in the past and there is little present guidance that can be offered with

respect to how these results should be judged. The true values of these calculations for diagnostic performance evaluation and stress testing will evolve as more experience is gained in their use and interpretation.

Analysis of Residuals. The set of residuals generated in a typical performance evaluation exercise contains significant information about contributions to "error" of: (a) the concentration data used for comparison with model output, (b) the soundness of the model formulation, and (c) the adequacy of the data supplied as input to the model. In addition to the operational evaluation procedures that are heavily based on the analysis of residuals, the evaluation may also profit from plotting residuals against selected variables in order to identify patterns of aberrant behavior. If correlations (relationships) between the residuals and one or more selected variables can be found, the emergent pattern (s) may be suggestive of the cause(s) of failure or inadequacy in the model. Variables that may be selected for plotting against residuals include time, geographical location, concentration levels, meteorological variables, emissions, and deposition rates. Plots can be made for the full region of interest and for the full duration of the simulation, or for subregions, selected time periods, and specified ranges in variables.

5.2.3 Longer Term Needs in Performance Evaluation

Testing Model Response to Emission Changes. The adequacy of a photochemical model in correctly predicting the effects of emission changes on ambient air quality should be evaluated directly, by examining model performance for applications involving significantly altered emission strengths and spatial patterns. Ideally, such a performance evaluation should be performed with emission inventories for the same region that correspond to years sufficiently apart. If such a procedure is not feasible, then, as an alternative it is suggested to evaluate the model for two or more different regions, using input data of comparable quality. The significant effects of meteorology on calculating the effects of emissions changes should be taken carefully into account in the evaluation process.

Emissions Model Testing. The emissions inventory input to a photochemical model is itself a model but it has received far less consideration in the modeling context despite the belief that the inventory may be the least accurate input in the photochemical modeling process. Estimation of uncertainties associated with emissions inventories is of considerable importance since any control strategy decisions made, with or without the aid of models, are only as good as the emissions estimates upon which they are based. New directions in the evaluation of emission inventory models include the following:

- Perform "Top down" vs. "Bottom up" Calculations. The emissions inventory should be developed for a pollutant, where possible, by estimating emissions of individual sources or groups of sources and then aggregating them and, independently, by calculating gross or integrated emissions rates.
- Perform Mass Balances. Mass balance calculations need to be considered as part of the inventorying process. For example, nitrogen balances can be made for determining emissions estimates and uncertainties for animal wastes, chemical fertilizer use, wastewater treatment plant emissions, and so on.
- Compare Ambient Air Ratios vs. Emissions Ratios. One should compare ratios of ambient concentrations of selected pollutant species, measured near the source, with estimated ratios of emissions rates of the two species. Analysis of differences in ratios may be used in estimating the magnitudes of emissions uncertainties.
- Consider Indirect Confirmations. External information should be used to the fullest extent possible to corroborate direct emissions estimates. For example, an energy balance on fuel consumption statistics has been used in estimating sulfur emission rates.

- Perform Source Testing. Emissions of selected sources should be determined through direct measurement. Because of the large attendant costs, one must develop specific guidelines for establishing this need and procedures for cost-benefit assessment.
- Carry Out Specially Designed Field Studies. Field studies should be considered specifically for the purpose of emissions determination. Examples include the SCAQS tunnel study (Ingalls et al., 1989) for corroborating estimates of emissions from vehicles.
- Improve Motor Vehicle Emissions Modeling. There is a clear need to improve the characterization of the driving cycle, including the degree of representativeness, extent of variability due to changes in commute characteristics, driver characteristics, and other influences.

Meteorological Model Testing. A wide range of meteorological models are available to support grid-based photochemical modeling. Prognostic meteorological models are becoming used more frequently, although the more sophisticated models are not necessarily superior to simpler models in all cases. If the observational network is exceptionally dense, interpolation techniques or simple diagnostic models may be cost-effective alternatives to the more complicated "primitive equation" models. Nevertheless, the prognostic meteorological models are expected to provide greater overall consistency in their predictions since they contain more of the relevant atmospheric physics. Eleven numerical measures for quantifying the performance of diagnostic and prognostic meteorological models are:

- Mean prediction;
- Mean observation;
- Standard deviation of predictions;

- Standard deviation of observations;
- Ratio of the predicted and observed standard deviations;
- Least squares regression statistics (slope and intercept);
- Root mean square error;
- Systematic root mean square error;
- Unsystematic root mean square error;
- Index of agreement; and
- Skill measures based on error and variance statistics.

Recommended graphical performance procedures include: (a) "slice plots", (b) surface windfields vector plots, (c) residual plots of hourly predicted and observed wind speeds, directions, mixing heights, and temperatures, and (e) plots of the mean wind speed, direction, wind speed standard deviation, the three root mean square errors, and the index of agreement as a function of simulation time.

Chemistry Model Testing. The chemistry module is one of the few components of photochemical models that has long been independently evaluated. Multi-species mechanism testing with environmental chamber data is the best available method of evaluating the chemistry for urban photochemical modeling. The recommended procedure for mechanism testing follows the hierarchy approach (Whitten 1983; Atkinson et al. 1987). Testing is initiated at the lowest level in the hierarchy (ie., NO_x-air experiments to test the inorganic reactions and the chamber characterization procedures) and proceeds to the highest level (>C₂ hydrocarbons) by step-wise addition of species with increasingly complex chemistry.

The research needs to improve mechanistic evaluation of chemistry models include: (1) acquiring chamber data from existing facilities and with existing methods for organic species for which there are little data; (2) developing and applying methods to collect data for species not presently measured in existing facilities; and (3) development of cleaner environmental chambers. The data collection and methods development efforts should focus on the following major areas of uncertainty in current photochemical mechanisms: the identification and subsequent reactions of aromatic ring-fragmentation

products, the radical yields in ozone-olefin reactions, the oxidation mechanisms of >C5 alkanes, and photolytic data for carbonyls.

Deposition Model Testing. Photochemical grid models include transport algorithms and chemical reaction resistance calculations to estimate dry deposition rates as a function of time and location. A theoretical evaluation of this methodology should be conducted to ensure that the procedures used agree with currently-accepted boundary layer theory and surface resistance experiments (see, for example, Hubbe, 1989). Field data should be sought on the atmospheric concentration and deposition of a pollutant that has essentially zero chemical reaction resistance to deposition. Since a theoretical test of the chemical reaction resistance algorithm in most deposition models is not practical, a literature survey the experimental data on the chemical reaction resistance for specific pollutant-surface pairs should be conducted. Then, sensitivity analysis procedures can be used to assess the uncertainty in the deposition module results due to the underlying uncertainty in the chemical reaction resistance values.

Treating Incommensurability. Photochemical grid models estimate average concentrations in each grid cell whereas measurements are made at a point. These different spatial (and temporal) scales of measurements and model predictions give rise to the so-called incommensurability problem. Potential solutions to this problem fall in several research areas:

- Sub-grid Scale Modeling. One may include in grid-based models those dynamic processes, such as chemical reaction, mixing, dispersion and deposition, that occur at spatial scales less than the size of the modeling grid.
- Finer Grid Resolution. One may reduce the dimensions of a grid cell size, although this leads to increased computing requirements and raises questions about how small the cell size can be reduced before the theoretical formulation of the model is compromised (Lamb, 1973).

- Remote Measurements. Multiple measurements employing remote sensing techniques may permit local concentrations to be compared directly with grid cell-averaged model predictions. Equipment potentially useful for this purpose is still in development.
- Multiple Measurements Within a Cell. Monitoring at multiple sites within a grid cell provides an alternate means for estimating grid-averaged concentrations, although current in-situ measurement costs make this approach impractical except for research purposes.

No widely applicable approach for resolving the commensurability problem is now available. The development of economical remote measurement systems and very low cost surface monitoring devices provide the most attractive routes to resolution.

Stress Testing. "Stressing a model" is designing and carrying out mechanistic tests that cause a model to reveal its flaws and weaknesses if it is indeed inadequate. Stressful testing is intended to reduce (or avoid) the risks associated with "weak" or otherwise inadequate tests, wherein a model is not challenged sufficiently to reveal its flaws and weaknesses or appears to be performing acceptably despite significant inadequacies in formulation or inputs. No prescription of "stress testing" now exists; however, certain requirements might reduce the risk of inappropriately accepting a flawed model for use in control strategy assessment. These include:

- Give equal, or nearly equal, weight to accurate estimation of NO₂ and volatile organic compounds as that given to ozone;
- Attempt, insofar as possible, to assure commensurateness in space and time of predictions and observations; that is, attempt to assure that observations are representative of predictions;

- Test the performance of individual model components, such as the chemical mechanism or deposition algorithm, prefatory to examining overall model performance, whenever possible;
- Test the performance of "preprocessor models", notably the meteorological and emissions models, whenever possible;
- Examine performance of the model over the range of meteorological, air quality (chemical), and emissions conditions available for study; and
- Require "hands off" testing after an adequate period of time is allowed for the "hands on" testing that is traditionally carried out during model development. Also, require testing by independent parties (who have no vested interest in the outcome) after the inception of the period of "hands off" testing.

Stress tests of a model should be "reasonable." A test should not attempt to demonstrate that a model cannot do what it was not designed to do and tests should not require data that cannot be acquired.

Use of Simplified Models. Simplified simulation models such as photochemical box models and trajectory models (including EKMA) should not be in general considered as alternatives to grid-based photochemical models. The potential usefulness of trajectory models is their utility in (1) performing a preliminary analysis of the application(s) relevant to which the grid-based model is going to be evaluated, to identify aspects of the modeling task that need to be studied in detail and therefore are appropriate as focal points of the full scale performance evaluation process, and (2) in quickly evaluating the effect of input data and parameter changes on pollutant levels, i.e., in performing efficient, low-cost, sensitivity analyses specifically relevant to the transformation and removal components of grid-based photochemical models. Box models can be used

to perform mass balances and check sensitivity to boundary and initial conditions for the entire region covered.

Model Modifications. Improved techniques are available for measuring nitric acid and aerosol nitrate as well as PAN in the field. The inorganic nitrate predictions from existing models are fairly inaccurate because they lump together two species that deposit at very different rates (i.e., nitric acid deposits very rapidly, while aerosol nitrate deposits slowly). Modifying the models to incorporate the nitric acid-aerosol nitrate equilibrium chemistry will make it possible to evaluate the nitric acid and aerosol nitrate predictions, as part of a multi-species evaluation. Comparison of the observed and predicted ratios of PAN to total nitrate and TIN to total nitrate should be made to assess possible biases in the VOC and NO_x inputs to the models.

6.0 BIBLIOGRAPHY

- Alcamo J., and J. Bartnicki (1987), "A Framework for Error Analysis of a Long Range Transport Model with Emphasis on Parameter Uncertainty," Atmos. Environment, 21:2121.
- Ames, J., S.R. Hayes, T.C. Meyers, and D.C. Whitney (1985). SAI Airshed Model Operations Manuals. Volume II--System's Manual. U.S. Environmental Protection Agency (EPA-600/8-85-007b).
- Ames, J., T.C. Myers, L.E. Reid, D.C. Whitney, S.H. Golding, S.R. Hayes, and S.D. Reynolds (1985). SAI Airshed Model Operations Manuals. Volume I--User's Manual, U.S. Environmental Protection Agency (EPA-600/8-85-007a).
- Anderson, G.E., et al. (1977). "Air Quality in the Denver Metropolitan Region: 1974-2000," EPA-908/1-77-002, U.S. Environmental Protection Agency, Research Triangle Park, NC.
- Anthes, R.A. (1983). "Regional Models of the Atmosphere in Middle Latitudes", Monthly Weather Review, Vol. 111, pp. 1306-1335.
- API (1989). "Detailed Analysis of Ozone SIPs in Seven Areas Selected for Retrospective Evaluation of Reasons for SIP Failure", American Petroleum Institute Report No. 4502, prepared by Pacific Environmental Services, Reston, VA and Systems Applications, Inc., San Rafael, CA.
- ARB (1989a). "Draft Technical Guidance Document: Photochemical Air Quality Modeling", Modeling and Meteorology Branch, Technical Support Division, California Air Resources Board, Sacramento, CA.
- ARB (1989b). "A Protocol for Photochemical Modeling of the South Coast Air Basin Using the South Coast Air Quality Study (SCAQS) Database", Control Strategy Modeling Section, Technical Support Division, California Air Resources Board, Sacramento, CA.
- Daly, C., M.A. Yocke, R.L. Dennis, N. Laulainen, P.M. Roth, and S. Seilkop (1988). "Protocol for Evaluating the Predictive Performance of the Regional Acidic Deposition Models RADM and ADOM", EPA-68-02-4129, Atmospheric Sciences Research Laboratory, Research Triangle Park, NC.
- Atkinson, R., H. Jeffries, G. Whitten, and F. Lurmann (eds.) (1987). Proceedings of Workshop on Evaluation/Documentation of Chemical Mechanisms, U.S. EPA, Atmospheric Sciences Research Laboratory, Research Triangle Park, NC (EPA/600/9-87/024).
- Atkinson, R., A.C. Lloyd, and L. Wings (1982). "An Updated Chemical Mechanism for Hydrocarbon/Nitrogen Oxides Photooxidations Suitable for Inclusion in Atmospheric Simulation Models." Atmospheric Environment, Vol. 16, pp. 1341-1355.
- Barchet, W.R. (ed.) (1987). "Evaluation of Regional-Scale Air Quality Models. Chairment Reports From Four Workshops." EPRI EA-5473, Project 1630-40. Final Report, October 1987.

- Beck, M.B. (1987). "Water Quality Modeling: A Review of the Analysis of Uncertainty," Water Resources Research, Vol. 23, No. 8, pp. 1393-1442.
- Bencala, K.E., and J.H. Seinfeld (1979). "Air Quality Model Performance Assessment Package," Atmospheric Environment, 13:1181-1185.
- Benkovitz, C.M., and N. L Oden (1989). "Individual Versus Averaged Estimates of Parameters Used in Large Scale Emissions Inventories," Atmospheric Environment, Vol. 23, No. 5, pp. 903-909.
- Blumenthal, D.L., J.G. Watson, and P.T. Roberts (1987). "Southern California Air Quality Study (SCAQS) - Program Plan," final report to California Air Resources Board by Sonoma Technology, Inc.
- Book, D.L., J.P. Boris and K. Hain (1975). "Flux Corrected Transport, II: Generalizations of the Method," J. Comput. Phys., 18:248-283.
- Boris, J.P., and D.L. Book (1973). "Flux Corrected Transport, I: SHASTA, And Algorithm That Works," J. Comput. Phys., 11:38-69.
- Brost, R.A. (1988). "The Sensitivity to Input Parameters of Atmospheric Concentrations Simulated by a Regional Chemical Model," J. Geophys. Res., 93D:2371.
- Builtjes, P.J.H., and S.D. Reynolds (1983). "Modeling the Effects of Emission Controls in the Netherlands," Environment International, 9:573-580.
- Builtjes, P.J.H., K.D. van den Hout, and S.D. Reynolds (1982). "Evaluation of Performance of Photochemical Dispersion Model in Practical Applications," 13th International Technical Meeting on Air Pollution Modeling and Application; 14-17 September, Ile des Embriez, France.
- Burton, C.S. (1988). Critical Review Discussion Paper: Ozone Air Quality Models, JAPCA, 38:1119.
- California Air Resources Board (1985). "Kern County Ozone Modeling for SIP Revision," Technical Services Division, Sacramento, CA.
- Carter, W.P.L. (1988). "Documentation of a Gas-Phase Photochemical Mechanism for Use in Airshed Modeling", Final Report to the California Air Resources Board, Contract No. A5-122-32.
- Carter, W.P.L. (1989). "A Detailed Mechanism for the Gas-Phase Atmospheric Reactions of Organic Compounds", Atmospheric Environment, (in press).
- Carter, W.P.L., R. Atkinson, A.M. Winer, and J.N. Pitts (1982). "An Experimental Investigation of Chamber-Dependent Radical Sources", Int. Journal Chemical Kinetics, Vol. 14, pg. 1071.
- Carter, W.P.L., A.C. Lloyd, J. L. Sprung, and J.N. Pitts (1979). "Computer Modeling of Smog Chamber Data: Progress in Validation of a Detailed Mechanism for the Photooxidation of Propene and n-Butane in Photochemical Smog". Int. Journal Chemical Kinetics, Vol. 11, pg. 45.

Carter, W.P.L., and F.W. Lurmann (1989). "Evaluation of the RADM Gas-Phase Chemical Mechanism", SAPRC Report to the U.S. EPA, Agreement CR-814558-01, Riverside, CA.

Carter, W.P.L., F.W. Lurmann, R. Atkinson, and A.C. Lloyd (1986). "Development and Testing of a Surrogate Species Chemical Mechanism", EPA/600/3-86/031, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Chameides, W.L., R.W. Lindsay, J. Richardson, and C.S. Kiang (1988). "The Role of Biogenic Hydrocarbon in Urban Photochemical Smog: Atlanta as a Case Study," Science, 241:1473-1475.

Chatwin, P.C., and P.J. Sullivan (1979). "Basic Structure of Diffusing Contaminant," in Mathematical Modeling of Turbulent Diffusion in the Environment, Academic Press, NY, pp. 3-32.

Cole, H.S., C.F. Newberry, W. Cox, G.K. Moss, D. Layland (1982). "Application of the Airshed Model for Ozone Control in St. Louis," 75th Annual Meeting of the Air Pollution Control Assn., New Orleans, LA, June 20-25, 1982.

Csanady, G.T. (1973). Turbulent Diffusion in the Environment, D. Reidel Publishing Co., Dordrecht, Holland.

Dabberdt, W.F., and W. Viezee (1987). "South Coast Central Coast Cooperative Aerometric Monitoring Program (SCCCAMP)," Bulletin of the American Meteorological Society, Vol. 68, pp. 1098-1110.

Dana, M.T., R.C. Easter, and J.M. Thorp (1984). "OSCAR Experiment High-Density Network Data Report: Event 4. 21-23 April 1981", Prepared for U. S. Environmental Protection Agency. PNL-5152, Pt 4, Pacific Northwest Laboratory, Richland, WA.

Davidson, A., and M. Hoggan (1984). "Air Quality Trends in the South Coast Air Basin," Planning Division Report, South Coast AQMD.

DeMandel, R.E., T.E. Perardi, and L.H. Robinson (1980). "Future Sensitivity of NO₂ to Hydrocarbon and NO_x Controls: Results of Photochemical Diffusion Modeling in the San Francisco Bay Area," Second Joint Conference on Applications of Air Pollution Meteorology, American Meteorology Society, 24-27 March, 1980, New Orleans, LA.

DeMandel, R.E., L.H. Robinson, J.S.L. Fong, and R.Y. Wada (1979). "Comparisons of EPA Rollback, Empirical/Kinetic, and Physico-chemical Oxidant Prediction Relationships in San Francisco Bay Area," Journal of Air Pollution Control Agcy., 30:157-162.

Demerjian, K.L., J. A. Kerr, and J. C. Calvert, (1974). "The Mechanism of Photochemical Smog Formation", Adv. in Environ. Sci. and Tech., Vol. 4, John Wiley and Sons, New York, NY.

Demerjian, K.L. (1985). "Quantifying Uncertainty in Long-Range Transport Models: Summary of AMS Workshop on Sources and Evaluation of Uncertainty in Long-Range Transport Models," Bulletin, American Meteorological Society, Vol. 66, No. 12, pp. 1533-1540.

Dennis, R.L. (1986). "Issues, Design and Interpretation of Performance Evaluations: Ensuring the Emperor has Clothes," in Air Pollution Modeling and its Application V, edited by C. de Wispelaere, F.A. Schiermeier, and N.V. Gillani. Plenum Press, New York, N.Y., pp. 411-424.

Dennis, R.L., and M.W. Downton (1984). "Evaluation of Urban Photochemical Models for Regulatory Use," Atmospheric Environment, 18:2055-2069.

Dennis, R.L., M.W. Downton, and R.S. Keil (1983). "Evaluation of Performance Measures for Urban Photochemical Model," EPA 450/4-83-021, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Derwent R., and O. Hov O. (1988). "Application of Sensitivity and Uncertainty Analysis Techniques to a Photochemical Ozone Model," J. Geophys. Res., 93D:5185.

Dickson, R.J., and A.D. Hobbs (1989). "Evaluation of Emission Inventory Uncertainty Estimation Procedures," 82nd Annual Meeting of the Air and Waste Management Association, 25-30 June, Anaheim, CA.

Dodge, M. (1977). "Combined use of Modeling Techniques and Smog Chamber Data to Derive Ozone-Precursor Relationships", In Proceedings of Inter. Conference on Photochemical Oxidants, Raleigh, NC.

Downton, M.W., and R.L. Dennis (1985). "Evaluation of Urban Air Quality Models for Regulatory Use: Refinement of an Approach," Journal of Climate and Applied Meteorology, 24:161-173.

Duewer, W.H., M.C. MacCracken, and J.J. Walton (1978). "The Livermore Regional Air Quality Model, II: Verification and Sample Applications to San Francisco Bay Area," Journal of Applied Meteorology, 17:273-311.

Dunker, A.M. (1980). "Response of Atmospheric Reaction-Transport Model to Changes in Input Functions," Atmospheric Environment, 14:671-679.

Dunker, A.M. (1984). "The decoupled Direct Method for Calculating Sensitivity Coefficients in Chemical Kinetics," J. Chem. Phys., 81:2385.

Dunker, A.M. (1986). "The Reduction and Parameterization of Chemical Mechanisms for Inclusion in Atmospheric Reaction-Transport Models," Atmos. Environment, 14:671.

Durham, J., et al. (1986). "Regional Eulerian Model Field Study and Evaluations: Proposed Management and Technical Approaches," Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC.

EPA (1979). "Procedures for the Preparation of Emissions Inventories for Volatile Organic Compounds, II: Emission Inventory Requirements for Photochemical Air Quality Simulation Models", EPA-450/4-79-018, U.S. Environmental Protection Agency, Research Triangle Park, NC.

EPA (1986). "Guideline on Air Quality Models (Rev.)," EPA-450/2-78-027R, U.S. Environmental Protection Agency, Research Triangle Park, NC.

EPA (1988). "Emission Inventory Requirements for Post-1987 Ozone State Implementation Plans", EPA-450/4-88-019, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Fosberg, M. A., W. E. Marlatt, and L. Krupnak (1976). "Estimating Airflow Patterns Over Complex Terrain", U.S. Forest Service, Rocky Mountain Forest and Range Experiment Station, Res. Paper RM-162, pp. 1-16, Fort Collins, CO.

Fox, D.G. (1981). "Judging Air Quality Model Performance," Bulletin, American Meteorological Society, 62:599-609.

Fox, D.G. (1984). "Uncertainty in Air Quality Modeling," Bull. Am. Met. Soc., 65:27.

Freeman, D.L., et al. (1986). "Method for Propagating Measurement Uncertainties Through Dispersion Models," Journal of Air Pollution Control Assn., 36:246-253.

Garrett, A. J., and F. G. Smith (1984). "Two-Dimensional Simulations of Drainage Winds and Diffusion Compared to Observations", Journal of Climate and Applied Meteorology, Vol. 23, pp. 598-610.

Gery, M.W., G.Z. Whitten, and J.P. Killus (1988). "Development and Testing of the CBM-IV for Urban and Regional Modeling", EPA/600/3-88/012, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Ghim, Y.S., and J.H. Seinfeld (1988). "Trajectory Models and the Deformation of Air Parcels," Atmos. Environment, Vol. 22, pg. 25.

Goodin, W.R., G.J. McRae, and J.H. Seinfeld (1980). "An Objective Analysis Technique for Constructing Three-Dimensional Urban Scale Wind Fields," Journal of Applied Meteorology, Vol. 19, pp. 98-108.

Haney, J.L., and T.N. Braverman (1985). "Evaluation and Application of Urban Airshed Model in Philadelphia Air Quality Control Region," EPA-450/4-85-003, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Haney, J.L., and C.S. Burton (1988). "UAM-CBM-IV-Calculated Effects of Emission Reductions on Ozone Levels in the Philadelphia Metropolitan Area," Systems Applications, Inc., San Rafael, CA (SYSAPP-88/117).

Haney, J.L., and C.S. Burton (1989). "Use of Urban Airshed Model to Assess Ozone Responses to Reductions in VOC and NO_x Emissions in Philadelphia AQCR", 82nd Annual Meeting of the Air and Waste Management Assn., 25-30 June, Anaheim, CA.

Haney, J.L., and D.R. Souten, T.W. Tesche, L.R. Chinkin, H. Hogo, and M.C. Dudik (1986). "Evaluation and Application of PARIS Photochemical Model in South Central Coast Air Basin," Report to U.S. Environmental Protection Agency by Systems Applications, Inc., San Rafael, CA.

Haney, J.L., T.W. Tesche, and J.P. Killus (1983). "Application of the Systems Applications Airshed Model to the Philadelphia Metropolitan Area: 19 July 1979 Ozone Episode, U.S. Environmental Protection Agency, Contract No. 68-02-3582, Systems Applications, Inc., San Rafael, CA.

Hayes, S.R. (1979). "Performance Measures and Standards for Air Quality Simulation Models," EPA-450/4-79-033, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Hayes, S.R., et al. (1988). "Evaluation of South Coast Air Basin Alternative Ozone Attainment Scenarios Using the Urban Airshed Model," Systems Applications Inc., report No. SYSAPP-88/195, Prepared for Western Oil and Gas Assn., Glendale, CA.

Hecht, T., J. H. Seinfeld, and M. Dodge. (1974). "Further Development of a Generalized Kinetic Mechanism for Photochemical Smog", Environmental Science and Technology, Vol. 8, pp. 327-339.

Hillyer, M.J., S.D. Reynolds, and P.M. Roth (1979). "Procedures for Evaluating the Performance of Air Quality Simulation Models," EPA-450/4-79-033, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Hogo, H., and M.A. Yocke (1986). "Analysis of PARIS Model Sensitivity and Performance Improvement for Cities Services--San Miguel Project Application," Systems Applications, Inc., San Rafael, CA (SYSAPP-86/022).

Hogo, H., and M.A. Yocke (1987). "Evaluation of a Photochemical Air Quality Model with Extensions to Calculate Aerosol Dynamics and Visibility," SYSAPP-87/017, Systems Applications, Inc., San Rafael, CA.

Hogo, H., et al. (1988). "Urban Airshed Model Performance Evaluation for 5-7 June 1985," Systems Applications, Inc., report No. SYSAPP-88/138, prepared for the South Coast Air Quality Management District, App. V-R, 1988 Rev.; draft Air Quality Management Plan.

Horie, Y. (1987). "Ozone Episode Representativeness Study for the South Coast Air Basin," Valley Research Corporation, for South Coast Air Quality Management District, El Monte, CA.

Huebert, B.J. and C.H. Robert (1985). "The Dry Deposition of Nitric Acid to Grass," Journal of Geophysical Research, Vol. 90, D1, pp. 2085-2090.

Iman, R.L., J.M. Davenport, and D.K. Zeigler (1980). "Latin Hypercube Sampling" (User's Guide), SAND-1472, Sandia National Laboratories, Albuquerque, NM.

Ingalls, M.N., L.R. Smith, and R.E. Kirksey (1989). "Measurement of Of-Road Vehicle Emission Factors in the California South Coast Air Basin: Volume 1, Regulated Emissions", Southwest Research Institute Report No. SwRI-1604, Prepared for the Coordinating Research Council, Inc., Atlanta, GA.

Irwin, J.S., et al. (1987). "Relating Error Bounds for Maximum Concentration Estimates to Diffusion Meteorology Uncertainty," Atmospheric Environment, Vol. 21, No. 9, pp. 1927-1937.

Jacob, D.J., S. Sillman, J.A. Logan, and S.C. Wofsy (1989). "Least Independent Variables Method for Simulation of Tropospheric Ozone," J. Geophys. Res., 94D:8497.

Jeffries, H.E., K.G. Sexton, and J. R. Arnold (1988). "Validation Testing of New Mechanism With Outdoor Chamber Data- Volume 2: Analysis of VOC Data for the CB4 and CAL Photochemical Mechanisms", University of North Carolina Report to the U.S. Environmental Protection Agency, Cooperative Agreement CR-813107, Chapel Hill, NC.

Jeffries, H.E., K.G. Sexton, and C.N. Salmi (1981). "the Effect of Chemistry and Meteorology on Ozone Control Calculations Using Simple Trajectory Models and the EKMA Procedure", EPA-45-/4-81-034. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Kersteter, S.L. (1988). "Procedures for the Preparation of Emission Inventories for Precursors of Ozone, 1", prepared for U.S. Environmental Protection Agency, Contract No. 68-02-4396, prepared by Alliance Technologies, Chapel Hill, NC.

Kessler, R., M.A. Yocke, and R.E. Morris (1987). "Designing a Wind Field Module," report No. SYSAPP-87/052, Systems Applications, Inc., San Rafael, CA.

Kessler, R.C., and S.G. Douglas (1989). "Numerical Simulation of Mesoscale Airflow in the South Central Coast Air Basin", Draft Final Report, No. SYSAPP-89/077, prepared for Sigma Research Corporation, by Systems Applications, Inc., San Rafael, CA.

Keyser, D. (1977). "Sensitivity Tests and a Real Data Forecast for the Planetary Boundary Layer Using a Mixed-Layer Model", M. S. Thesis, Department of Meteorology, The Pennsylvania State University, 187 pp.

Keyser, D., and R.A. Anthes (1977). "The Applicability of a Mixed-Layer Model of the Planetary Boundary Layer to Real-Data Forecasting," Mon. Weather Review, Vol. 105, pg. 1351.

Killus, J.P., W.R. Oliver, P.D. Gutfreund, J.E. Langstaff, T.W. Tesche, and H.J. Su (1981). "Simulation of Impacts of Nitrogen Oxide Control Strategies Under Oxidant Episode Conditions," Report No. SYSAPP-81/165, Systems Applications, Inc., San Rafael, CA.

Killus, J.P. and G.Z. Whitten (1982). "A New Carbon-Bond Mechanism for Air Quality Modeling", EPA-600/3-82-041, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Killus, J.P., et al. (1977). "Continued Research in Mesoscale Air Pollution Simulation Modeling, IV: Refinements in Numerical Analysis, Transport, Chemistry, and Pollutant Removal," Report 77-142, Systems Applications, Inc., San Rafael, CA.

Lamb, R.G. (1973). 'Note on the Application of K-Theory to Diffusion Problems Involving Nonlinear Chemical Reactions,' Atmospheric Environment, Vol. 7, pp. 257-263.

Lamb, R.G. (1986). "Basic Problems and Future Solutions Over Large Time and Space Domains," Proceedings, 5th Joint Conference on Applications of Air Pollution Meteorology, American Meteorological Society, 18-21 November 1986, Chapel Hill, NC.

Lamb, R.G. (1988). "Simulated Effects of Hydrocarbon Emissions Controls on Seasonal Ozone Levels in Northeastern United States: A Preliminary Study," prepared by Atmospheric Sciences Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Lamb, R.G., and S.K. Hati (1987). "Representation of Atmospheric Motion in Models of Regional-Scale Air Pollution," Journal of Climate and Applied Meteorology, 26:837-846.

Langstaff, J.E. and J.K. Wagner (1986). "Estimation of Uncertainty for the 1980 NAPAP Emission Inventory," EPA-600/7-86-055, U.S. Environmental Protection Agency.

Lawson, D.R. (1990). "The Southern California Air Quality Study," Journal of the Air and Waste Management Association, Vol. 40, No. 2, pp. 156-165.

Layland, D.E. (1980). "Guideline for Applying Airshed Model to Urban Areas," EPA-450/4-80-020, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Layland, D.E., and H.S. Cole (1983). "Review of Recent Applications of SAI Urban Airshed Model," EPA-450/4-84-004, Office of Air Quality Planning and Standards, Environmental Protection Agency, Research Triangle Park, NC.

Layland, D.E., S.D. Reynolds, H. Hogo, and W.R. Oliver (1983). "Demonstration of Photochemical Grid Model Usage for Ozone Control Assessment," 76th Annual Meeting of the Air Pollution Control Assn., Atlanta, GA., 19-24 June.

Leone, J.A. and J.H. Seinfeld (1975). "Evaluation of Chemical Reaction Mechanisms for Photochemical Smog: Part II- Quantitative Evaluation of the Mechanisms". EPA-600/3-84-063. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Liu, M.K., T.C. Myers, and J.L. McElroy (1979). "Numerical Modeling of Land and Sea Breeze Circulation Along a Complex Coastline", Mathematics and Computers in Simulation, Vol. 21, No. 4, pp. 359-367.

Liu, M.K., and J.H. Seinfeld (1975). "On the Validity of Grid and Trajectory Models of Urban Air Pollution," Atmos. Environment, Vol. 9, pg. 555.

Ludwig, F.L., and E. Shelar (1978). "Site Selection for the Monitoring of Photochemical Air Pollutants," U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC 27711 (EPA-450/3-78-013).

Lurmann, F.W., W.P.L. Carter, and A.C. Lloyd (1987). "Surrogate Species Chemical Reaction Mechanism for Urban-Scale Air Quality Simulation Models," EPA/600/3-87/014, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Lurmann, F.L., A.C. Lloyd, and R. Atkinson (1986). "A Chemical Mechanism for Use in Long-Range Transport/Acid Deposition Computer Modeling", Journal of Geophysical Research, Vol. 91, No. D10, pg. 10,905-10,936.

Lurmann, F.W., A.C. Lloyd, and R.A. Atkinson (1986). "Suggested Gas-Phase Species Measurements for the Southern California Air Quality Study," in Analysis of the Southern California Air Quality Study (SCAQS) Measurement Plan, Coordinating Research Council, Atlanta, GA.

Mahrer, Y., and R. A. Pielke (1978). "A Test of an Upstream Spline Interpolation Technique for the Advection Terms in a Numerical Mesoscale Model", Monthly Weather Review, Vol. 106, pp. 818-830.

Martin, C. (1981). "Numerical Accuracy in a Mesoscale Meteorological Model", M.S. Thesis, University of Virginia, Department of Meteorology, Charlottesville, VA.

McNally, D.E., and T.W. Tesche (1988). "Contribution of OCS Emissions to Ozone Air Quality in South Central Coast Air Basin," report to U.S. Environmental Protection Agency, by Radian Corporation, Sacramento, CA.

McNally, D.E., and T.W. Tesche (1989). "Ventura Ozone Federal Implementation Plan." EPA Contract No. 68-02-4392-37 by Radian Corporation, Sacramento, CA.

McNally, D.E., T.W. Tesche, and W.R. Oliver (1989). "Ozone Control Implications Revealed by Airshed Model Isopleths," 82nd Annual Meeting of Air Pollution Control Assn., 25-30 June, Anaheim, CA.

McNider, R. T. (1981). "Investigation of the Impact of Topographic Circulations on the Transport and Dispersion of Air Pollutants", Ph.D. Dissertation, University of Virginia, Department of Environmental Sciences, Charlottesville, VA.

McRae, G.J., W.R. Goodin, and J.H. Seinfeld (1982). "Mathematical Modeling of Photochemical Air Pollution," EQL Report No. 18, CalTech, Pasadena, CA.

McRae, G.J., F.L. Lurmann, and T.W. Tesche (1988). "Review of Photochemical Air Quality Modeling Strategies Carried Out In Support of Draft Air Quality Management Plan," technical memorandum to South Coast Air Quality Management District, 19 September.

McRae, G.J., and A.G. Russell (1984). "Dry Deposition of Nitrogen Containing Species," in Hicks, B.B. editor, Deposition Both Wet and Dry, Butterworth Publishers, Boston.

McRae, G.J., and J.H. Seinfeld (1983). "Development of a Second Generation Mathematical Model for Urban Air Pollution, II: Evaluation of Model Performance," Atmospheric Environment, 17:501.

Mayerson, H.J., et al. (1975, 1976). "Atmospheric Hydrocarbon Concentrations," California Air Resources Board, Sacramento, CA.

McRae, G.J. (1988). "Role of Hydrocarbons and NO_x," Conference on Photochemical Modeling as a Tool for Decision Makers, California Air Resources Board, 1-3 February, Pasadena, CA.

Milford, J.B., A.G. Russell, and G.J. McRae (1989). "A New Approach to Photochemical Pollution Control: Implications of Spatial Patterns in Pollutant Responses to Reductions in Nitrogen Oxides and Reactive Organic Gas Emissions," Environmental Science and Technology, Vol. 23, No. 10, pp. 1290-1301.

Moore, G.E., C. Daly, and M. K. Liu (1987). "Modeling of Mountain-Valley Wind Fields in the Southern San Joaquin Valley, California", Journal of Climate and Applied Meteorology, Vol. 26, pp. 1230-1242.

Moore, G.E., T.E. Stoeckenius, and D.A. Stewart (1982). "A Survey of Statistical Measures of Model Performance and Accuracy for Several Air Quality Models," Report to the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Systems Applications, Inc., No. EF 82/072.

Morris, R.E. (1989). "Urban Airshed Model Studies of Five Cities. Low-Cost Application of the Model to Future Year SIP Control and Alternative Fuel Strategies for Dallas-Fort Worth, Atlanta, Philadelphia, and St. Louis," Systems Application, Inc., San Rafael, CA.

Morris, R.E., L. Chinkin, L. Gardner, L. Mahoney, J. Chamberlin, R. Miles-McLean, and R. Scheffe (1989). "Use of Urban Airshed Model to Assess Effects of Ethanol-Blended Fuels on Ozone Concentrations in New York and St. Louis", 82nd Annual Meeting of the Air and Waste Management Assn., 25-30 June, Anaheim, CA.

Morris, R.E., R.C. Kessler, S.G. Douglas, K.R. Styles, G.E. Moore (1988). "Rocky Mountain Acid Deposition Model Assessment Acid Rain Mountain Mesoscale Model (ARM3)." Systems Applications, Inc., San Rafael, California (SYSAPP-88/152).

Morris, R.E., T.C. Myers, E.L. Carr, and M.C. Causley (1989d). "Urban Airshed Model Study of Five Cities. Demonstration of Low-Cost Application of the Model to the City of Atlanta and the Dallas-Fort Worth Metroplex Regions," Systems Application, Inc., San Rafael, CA (SYSAPP-89/122).

Morris, R.E., T.C. Myers, M.C. Causley, L. Gardner, and E.L. Carr (1989). "Urban Airshed Model Study of Five Cities. Low-Cost Application of the Model to Atlanta and Evaluation of the Effects of Biogenic Emissions on Emission Control Strategies," Systems Applications, Inc., San Rafael, CA (SYSAPP-89/117).

Morris, R.E., T.C. Myers, H. Hogo, L.R. Chinkin, L.A. Gardner, and R.G. Johnson (1989). "A Low-Cost Application of the Urban Airshed Model to the New York Metropolitan Area and the City of St. Louis," Systems Applications, Inc., San Rafael, CA (SYSAPP-89/038).

Morris, R.W., M.W. Gery, M.K Liu, G.E. Moore, C. Daly, and S.M. Greenfield (1989). "Sensitivity of a Regional Oxidant Model to Variations in Climate Parameters. Volume I: Results," Systems Applications, Inc., San Rafael, California (SYSAPP-89/014a).

Morris, R.W., T.C. Myers, and E.L. Carr (1989). "Urban Airshed Model Study of Five Cities: Evaluation of Base Case Model Performance for the City of St. Louis Using Rich and Sparse Meteorological Inputs," Systems Applications, Inc., San Rafael, CA.

Munn, R.E., J. Alcamo, and V. Fedorov (1988). "Evaluating the Performance of Air Quality Models in a Policy Framework," in Air Pollution Modeling and its Application VI, edited by H. van Dop. Plenum Press, New York, N.Y. pp.237-257.

Myers, T.C., H. Hogo, R.G. Johnson, R. Scheffe, J. Chamberlin, and R. Miles-McLean (1989). "Evaluation of Low Cost Application of Urban Airshed Model for City of St. Louis," 82nd Annual Meeting of the Air and Waste Management Assn., 25-30 June, Anaheim, CA.

Myers, T.C., D.C. Whitney, and S.D. Reynolds (1979). "Photochemical Air Quality Simulation Modeling: A Manual for Transportation Planners and Highway Engineers, Volume II-Appendix E. User's Guide to the SAI Trajectory Model," Systems Application, Inc., San Rafael, CA (SYSAPP-79/003R).

Oliver, W.R., H. Hogo, and P. Saxena (1983). "Assessment of NO_x Emission Control Requirements in California South Coast Air Basin, IV, App. C: Emission Inventory Review and Development," prepared for Western Oil and Gas Assn. by Systems Applications, Inc., San Rafael, CA.

Oliver, W.R., and T.W. Tesche (1986). "Are Today's Speciated Hydrocarbon Inventories Sufficient for Tomorrow's Acid Deposition and Regional Oxidant Program Needs?" presented at 79th Annual Meeting Air Pollution Control Assn., June 22-27, Minneapolis, MN.

Oliver, W.R., T.W. Tesche, S.H. Peoples, R.R. Boyd, and R. Tate (1987). "Planning Study for Emission Inventory," report to Emissions Working Group, Southern California Air Quality Study, DCN-87-264-013-02, Radian Corporation, Sacramento, CA.

O'Neill, R.V., R.H. Gardner, and J.H. Carney (1982). "Parameter Constraints in a Stream Ecosystem Model: Incorporation of a priori Information in Monte Carlo Error Analysis," Ecological Modeling, 16:51.

Pielke, R.A. (1984). Mesoscale Meteorological Modeling, Academic Press.

Pielke, R.A. (1985). "The Use of Mesoscale Numerical Models to Assess Wind Distribution and Boundary-Layer Structure in Complex Terrain", Boundary Layer Meteorology, Vol. 31, pp. 217-231.

Pielke, R.A., and Y. Mahrer (1978). "Verification Analysis of the University of Virginia Three-Dimensional Mesoscale Model Prediction Over South Florida", Monthly Weather Review, Vol. 106, pp. 1568-1589.

Pierson, W.R., W.W. Brachaczek, S.M. Japar, G.R. Cass, and P.A. Solomon (1988). "Dry Deposition and Dew Chemistry in Claremont, California, During the 1985 Nitrogen Species Methods Comparison Study," Atmospheric Environment, Vol. 22, pp. 1657-1663.

Pilinis, C. and J.H. Seinfeld (1988). "Development and Evaluation of Eulerian Photochemical Gas-aerosol Model", Atmospheric Environment, Vol. 22, No. 9, pp. 1985-2001.

Pollack, A.K., T.E. Stoeckenius, J.L. Haney, T.S. Stocking, J.L. Fieber, and M. Moezzi (1988). "Analysis of Historical Ozone Concentrations in the Northeast," System Applications Inc., San Rafael, CA (SYSAPP-88/192).

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling (1986). Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, London.

Preisendorfer, R.W., and T.P. Barnett (1983). "Numerical Model-Reality Intercomparison Tests Using Small-Sample Statistics", Journal of the Atmospheric Sciences, Vol. 40, pp. 1884-1896.

Radian (1987). "Planning Study for SCAQS Emission Inventory: Recommendations for Collection of Day-Specific Emission-Related Data", prepared for Emissions Working Group, Southern California Air Quality Study, by Radian Corporation, Sacramento, CA.

Rao, S.T., et al. (1987). "Application of Urban Airshed Model to New York Metropolitan Area," North American Oxidant Symposium, 25-27 February, Quebec City, Canada.

Ray, P.S., 1986. Mesoscale Meteorology and Forecasting, edited by P. S. Ray, American Meteorological Society, Boston, Mass.

Reynolds, S.D., L. Reid, M.I. Hillyer, J.P. Killus, T.W. Tesche, R.I. Pollack, and G.E. Anderson (1979). "Photochemical Modeling of Transportation Control, I: Model Development, Performance Evaluation, and Strategy Assessment," report to Federal Highway Department, FHWA-RD-78-173.

Reynolds, S.D., T.W. Tesche, and L.E. Reid (1979). "Introduction to the SAI Airshed Model and Its Usage", SYSAPP-79-31, Systems Applications, Inc., San Rafael, CA.

Reynolds, S.D., T.W. Tesche, D.R. Souten, and M. Liu (1985). "Overall Study Protocol for the South Central Coast Cooperative Aerometric Monitoring Program," final report to Western Oil and Gas Association by Systems Applications, Inc.

Reynolds, S.D., et al. (1974). "Mathematical Modeling of Photochemical Air Pollution, III: Evaluation of the Model," Atmospheric Environment, 8:563-596.

Reynolds, S.D., et al. (1976). "Continued Development and Validation of 2nd Generation Photochemical Air Quality Simulation Model, II: Refinements in Treatment of Chemistry, Meteorology, and Numerical Integration Procedures," EF75-24R, EPA-600/4-26-016b, Systems Applications Inc., San Rafael, CA.

Reynolds, S.D., et al. (1979). "Photochemical Modeling of Transportation Control Strategies," report no. EF79-37R, Systems Applications Inc., San Rafael, CA.

Reynolds, S.D., et al. (1982). "Application of SAI Airshed Model to Tulsa Metropolitan Area," report No. SYSAPP-82/004, Systems Applications Inc., San Rafael, CA.

Riggan, P.J., R.N. Lockwood, and E.N. Lopez (1985). "Deposition and Processing of Airborne Nitrogen Pollutants in Mediterranean-Type Ecosystems of Southern California," Environmental Science and Technology, Vol. 19, pp. 781-789.

Roth, P.M. (1988). "History of Ozone Modeling," presentation at Conference on Photochemical Modeling as a Tool for Decision Makers, Co-sponsored by the California Air Resources Board and CalTech, 1-3 February 1988, Pasadena, CA.

Roth, P.M., C.L. Blanchard, and S.D. Reynolds (1989), "The Role of Grid-Based, Reactive Air Quality Modeling in Policy Analysis: Perspectives and Implications, As Drawn From a Case Study", Report to the U.S. Environmental Protection Agency, Atmospheric Sciences Modeling Division, Contract No. 68-01-6849, Research Triangle Park, NC.

Roth, P.M., P.G. Georgopoulos, T.B. Smith, A.E. Eschenroeder, J.H. Seinfeld, P.H. Guldberg, and T.C. Spangler (1988). "Guidelines for Air Quality Modeling", Final Report to the California Air Resources Board.

Roth, P.M., S.D. Reynolds, T.W. Tesche, P.D. Gutfreund, and C. Seigneur (1983). "Appraisal of Emissions Control Requirements in California South Coast Air Basin," Environment International, 9:549-571.

Roth, P.M., C. Seigneur, T.W. Tesche, and S. D. Reynolds (1983b). "Assessment of NO_x Emission Control Requirements in California South Coast Air Basin, III: Results, Findings, and Implications for Control Requirements Based on Short-term Modeling Studies", report to Western Oil and Gas Assn., by Systems Applications, Inc., San Rafael, CA.

Ruff, R.E. (1983). "Application of Statistical Methods to Disguise Causes of Poor Air Quality Model Performance," Atmospheric Environment, 17:291-297.

Russell, A.G., and G.R. Cass (1986). "Verification of Mathematical Model for Aerosol Nitrate and Nitric Acid Formation and Use for Control Measure Evaluation," Atmospheric Environment, Vol. 20, No. 10, pp. 2011-2055.

Russell, A.G., K.F. McCue, G.R. Cass (1988). "Mathematical Modeling of Formation of Nitrogen-Containing Air Pollutants, I: Evaluation of an Eulerian Photochemical Model," Environmental Science and Technology, 22:263.

Russell, A.G., K.F. McCue, and G.R. Cass (1988). "Mathematical Modeling of Nitrogen-containing Pollutants, 2: Evaluation of Effects of Emission Controls," Environmental Science and Technology, 22:1336-1347.

SAI (1989). "User's Manual for Preparing Emission Files for Use in the Urban Airshed Model", Systems Applications, Inc., San Rafael, CA (SYSAPP-89/114).

Schere, K. (1988). "Ozone-Air Quality Models: Critical Review Discussion Papers," JAPCA, Vol. 38, pp. 1114.

Schere, K.L. (1983). "Evaluation of Several Numerical Advection Schemes," Atmospheric Environment, Vol. 17, No. 10, pp. 1987-1907.

Schere, K.L., and J.H. Shreffler (1982). "Final Evaluation of Urban Scale Photochemical Air Quality Simulation Models," EPA-600/3-82-094, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Segal, M., Y. Mahrer, and R. A. Pielke (1982b). "Application of a Numerical Mesoscale Model for the Evaluation of Seasonal Persistent Regional Climatological Patterns", Journal of Applied Meteorology, Vol. 21, pp. 1754-1762.

Segal, M., R. T. McNider, R. A. Pielke, and D. S. McDougal (1982a). "A Numerical Model Simulation of the Regional Air Pollution Meteorology of the Greater Chesapeake Bay Area-Summer Day Case Study", Atmospheric Environment, Vol. 16, No. 6. pp1381-1397.

Seigneur, C.S., T.W. Tesche, P.M. Roth, and L.E. Reid (1981). "Sensitivity of Complex Air Quality Model to Input Data", Journal of Applied Meteorology, Vol. 20, No. 9.

Seigneur, C.S., et al. (1983). "On the Treatment of Point Source Emissions in Urban Air Quality Modeling," Atmospheric Environment, Vol. 17, No. 9, pp. 1655-1676.

Seinfeld, J.H. (1986). Atmospheric Chemistry and Physics of Air Pollution, John Wiley and Sons, New York, NY.

Seinfeld, J.H. (1988a). "Ozone Air Quality Models: A Critical Review," Journal of Air Pollution Control Assn., Vol. 38, No. 5, pp. 616-645.

Seinfeld, J.H. (1988b). "State-of-the-Art for Ozone Modeling", presentation at Conference on Photochemical Modeling as a Tool for Decision Makers, co-sponsored by the California Air Resources Board and the California Institute of Technology, 1-3 February, Pasadena, CA.

Seinfeld, J.H. (1989). "Urban Air Pollution: State of the Science," Science, Vol. 243, pp. 745-752.

Seinfeld, J.H., R. Atkinson, G.R. Cass, K.T. Knapp, T.B. Smith, T.W. Tesche, and J.G. Watson (1988). "Recommendations on Analysis of Data from Southern California Air Quality Study (SCAQS)", report prepared by SCAQS Data Analysis Committee for California Air Resources Board, 15 April, 1988.

Seinfeld, J.H., D. Blumenthal, J. Carroll, G. Cass, J. Hales, M. Lev-On, F.W. Lurmann, W.R. Oliver, and T.W. Tesche (1987). "Urban Airshed Model Technical Advisory Committee Recommendations", report to Western Oil and Gas Assn. and the California Air Resources Board, 15 April, 1987.

Shareef, G.S., W.A. Butler, L.A. Bravo, and M.B. Stockton (1988). "Air Emissions Species Manual, 1: Volatile Organic Compound Species Profiles", report to U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, EPA-450/2-88-003a, prepared by Radian Corp., Research Triangle Park, NC.

Simpson, J., N.E. Westcott, R.J. Clerman, and R.A. Pielke (1979). "On Cumulus Mergers", Arch. Met. Geophys. Bioklim., Vol. A29, pp. 1-40.

SLC (1986). "Draft Environmental Impact Report/Environmental Impact Statement for Proposed ARCO Coal Oil Point Project," Prepared by the California State Lands Commission, Sacramento, CA.

Steyn, D. G., and I.G. McKendry (1988). "Quantitative and Qualitative Evaluation of a Three-Dimensional Mesoscale Numerical Model Simulation of a Sea Breeze in Complex Terrain", Monthly Weather Review, Vol. 116, pp. 1914-1926.

Stoeckenius, T.E. (1989). "Comparison of Techniques for Adjusting Ozone Exceedances to Account for Variations in Meteorological Conditions," paper presented at the 82nd Annual Meeting of the Air and Waste Management Association, June 25-30, 1989, Anaheim, CA.

Stoeckenius, T.E., W.M. Cox, J.A. Tikvert, and J. Pearson (1985). "Method for Calculating Dispersion Modeling Uncertainty Applied to Regulation of Emission Source," Proceedings, 78th Annual Air Pollution Control Assn., June 1985, Detroit, MI.

Tennekes, H. (1988). "The Outlook: Scattered Showers", Bulletin of the American Meteorological Society, Vol. 69, No. 4, pp. 368-372.

Tesche, T.W. (1979). "Summary of Hypothetical 1982 Simulations Examining the Influence of Power Plant NOx Controls of Photochemical Oxidants in South Coast Air Basin," prepared for Southern California Edison Co., by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W. (1981a). "Compilation of Maximum Ozone and NO2 Deficit-Enhancement Plots for 18 Airshed Simulations, 26-27 June 1974," report to the California Air Resources Board, by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W. (1981b). "Sensitivity of Complex Photochemical Model Estimates to Detail in Input Information. App. B: Specification and Assessment of Airshed Model Input Requirements", report to U.S. Environmental Protection Agency, by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W. (1983). "Assessment of NOx Emission Control Requirements in the California South Coast Air Basin: Performance Evaluation of Airshed Model for 7-8 November 1978 NO2 Episode," SYSAPP-82/230, Systems Applications, Inc., San Rafael, CA.

Tesche, T.W. (1983). "Photochemical Dispersion Modeling: Review of Model Concepts and Applications Studies," Environment International, 9:465-489.

Tesche, T.W. (1986). "Comparison of Two Numerical Schemes for Integrating Atmospheric Diffusion Equation: Evaluation with Atmospheric Data", International Journal of Mathematical Modeling, Vol. 9, No. 7, pp. 507-519.

Tesche, T.W. (1987a). "Comparison of Two Numerical Schemes for Integrating Atmospheric Diffusion Equations: Evaluation With Atmospheric Data," International Journal of Mathematical Modeling, Vol. 9, No. 7, pp. 507-519.

Tesche, T.W. (1987). "Examination of the Underprediction Problem in California OCS Photochemical Modeling Studies: Causes and Potential Solutions," 80th Annual Meeting of the Air Pollution Control Assn., 21-26 June, New York, NY.

Tesche, T.W. (1988). "Accuracy of Ozone Air Quality Models", Journal of Environmental Engineering, Vol. 114, No. 4, pp. 739-752.

Tesche, T.W. (1988b). "Non-Traditional Uses of the Airshed Model in Quantifying Expected Benefits of Urban Emissions Control Strategies", presented at the 81st Annual Meeting of the Air Pollution Control Assn., 21-25 June, Dallas, TX.

Tesche, T.W. (1989a). "Urban Airshed Model Sensitivity/Uncertainty: Review of 15 Years of Simulation Experience", report to Houston Regional Monitoring Corporation, by Radian Corporation, Sacramento, CA.

Tesche, T.W. (1989b). "Improvement of Procedures for Evaluating Photochemical Models: Strawman Plan", report to Research Division, California Air Resources Board, by Radian Corporation, Sacramento, CA.

Tesche, T.W., and C.S. Burton (1978). "Simulated Impact of Alternative Emission Control Strategies on Photochemical Oxidants in Los Angeles, II: Comparisons of Observed and Predicted Concentrations," No. EF78-3, by Systems Applications, Inc., San Rafael, CA

Tesche, T.W., C.S. Burton, and V.A. Mirabella (1979). "Recent Verification Studies with SAI Urban Airshed Model in South Coast Air Basin", 4th Symposium on Turbulence, Diffusion, and Air Pollution, American Meteorological Society, 15-18 January 1979, Reno, NV.

Tesche, T.W., and J.L. Haney (1981). "Comparison Between Airshed and PARIS Model Oxidant Predictions for South Coast Air Basin," report to Electric Power Research Institute, no. SYSAPP-1297, by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., J.L. Haney, and R.E. Morris (1987a). "Performance Evaluation of Four Grid-Based Dispersion Models in Complex Terrain," Atmospheric Environment, Vol. 21, No. 1, pp. 233-256.

Tesche, T.W., J.L. Haney, and D.R. Souten (1986). "Application of Advanced Photochemical Models to Offshore Petroleum Resource Development," 79th Annual Meeting of the Air Pollution Control Assn., June 26-27, Minneapolis, MN.

Tesche, T.W., and D.E. McNally (1989). "A Three-Dimensional Photochemical-Aerosol Model for Episodic and Long-Term Simulation: Formulation and Initial Application in the Los Angeles Basin," Proceedings of the American Chemical Society Symposium on Atmospheric Modeling, Miami Beach, FL.

Tesche, T.W. and D.E. McNally (1989a). "Urban Airshed Modeling in Support of Ventura Ozone Federal Implementation Plan," report to U.S. Environmental Protection Agency, Region IX, by Radian Corporation, Sacramento, CA.

Tesche, T.W. and D.E. McNally (1989b). "Photochemical Modeling of Texaco's Proposed Offshore Exploration Activities on Leases P-0514-17", prepared for ERC, Environmental Services, by Radian Corporation, Sacramento, CA.

Tesche, T.W., D.E. McNally, and W.R. Oliver (1988b). "Airshed Model Simulation of 6-7 September 1984 Ozone Episode in Support of Santa Barbara Air Quality Attainment Plan," report to Santa Barbara County Air Pollution Control District by Radian Corporation, Sacramento, CA.

Tesche, T.W., D.E. McNally, and W.R. Oliver (1989). "Ozone Control Implications Revealed by Airshed Model Isopleths", Proceedings, 82nd Annual Meeting of the Air Pollution Control Assn., 25-30 June, Anaheim, CA.

Tesche, T.W., D.E. McNally, and J.G. Wilkinson (1988). "Importance of Boundary Layer Measurements for Urban Airshed Modeling," 81st Annual Meeting of Air Pollution Control Assn., 20-24 June, Dallas, TX.

Tesche, T.W., T.C. Myers, C. Daly, and L.R. Chinkin (1985). "Cumulative Ozone Impact Assessment for Kern River Cogeneration Company's Sycamore Project," no. SYSAPP-85/156, Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., T.C. Myers, and G.E. Moore (1986). "Use of Dynamic Meteorological Models for Input to Photochemical Dispersion Models", Proceedings, 79th Annual Air Pollution Control Assn. Meeting, 22-27 June 1986, Minneapolis, MN.

Tesche, T.W., and W.R. Oliver (1986). "Decade of Experience in Photochemical Grid Modeling: Synthesis of U.S. and European Model Evaluation Studies," 5th Joint Conference on Applications of Air Pollution Meteorology, American Meteorological Society/Air Pollution Control Assn., Chapel Hill, NC.

Tesche, T.W., W.R. Oliver, and J.L. Haney (1981). "Sensitivity of Urban Airshed Model Ozone Predictions to Emission Inventory Perturbations," report to South Coast Air Quality Management District, by Systems Applications, Inc., No. 81225, San Rafael, CA.

Tesche, T.W., W.R. Oliver, and J.L. Haney (1982). "Grid-based Photochemical Dispersion Modeling of Proposed SCE Ivanpah Generating Station," prepared for Southern California Edison Co., by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., W.R. Oliver, H. Hogo, P. Saxena, and J.L. Haney. (1982a). "Assessment of NO_x Emission Control Requirements in California South Coast Air Basin: Performance Evaluation of Airshed Model for 26-27 June 1974 Ozone Episode," No. SYSAPP-82/064, Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., W.R. Oliver, H. Hogo, P. Saxena, and J.L. Haney. (1982b). "Assessment of NO_x Emission Control Requirements in California South Coast Air Basin: Performance Evaluation of Airshed Model for 7-8 November 1978 NO₂ Episode," No. SYSAPP-82/230, Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., W.R. Oliver, H. Hogo, P. Saxena, and J.L. Haney (1983a). "Assessment of NO_x Emission Control Requirements in California South Coast Air Basin. IV: Performance Evaluation of Systems Applications Airshed Model for 26-27 June 1974 O₃ Episode in South Coast Air Basin", no. SYSAPP-83/037, Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., W.R. Oliver, H. Hogo, and D.R. Souten (1980). "Simulated Ozone Concentrations in South Coast Air Basin: Comparison Between 1974 and 1987 Levels," report to National Commission on Air Quality, no. 244-ET80-143R, by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., and P.M. Roth (1988). "Performance Evaluation: Needs and Future Directions", presentation at the Conference on Photochemical Modeling as Tool for Decision Makers, Co-sponsored by California Air Resources Board and CalTech, 1-3 February, 1988, Pasadena, CA.

Tesche, T.W., C. Seigneur, W.R. Oliver, J.C. Haney (1982). "Numerical Modeling of Photochemical Smog in Los Angeles: Examination of Alternative Emission Reduction Strategies", report to the South Coast Air Quality Management District, no. 83032, by Systems Applications, Inc., San Rafael, CA.

Tesche, T.W., C. Seigneur, L. E. Reid, P.M. Roth, W. R. Oliver, and J.C. Casmassi, (1981). "Sensitivity of Complex Photochemical Model Estimates to Detail in Input Information," No. SYSAPP-81/5, prepared for the U.S. Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

Tesche, T.W., C. Seigneur, L.E. Reid, P.M. Roth, W.R. Oliver, and J.C. Cassmassi (1981). "The Sensitivity of Complex Photochemical Model Estimates to Detail in Input Information," EPA-450/4-81-031a, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Tesche, T.W., J.G. Wilkinson, D.E. McNally (1988a). "Airshed Modeling of San Francisco Bay Region: Evaluation of Ozone Impact of Proposed Crockett Cogeneration Facility under Four Different Meteorological Scenarios", report to California Energy Commission, by Radian Corporation, Sacramento, CA.

Tesche, T.W., J.G. Wilkinson, and D.E. McNally (1988). "Role of Wind Field Estimations in Airshed Modeling for Ozone Non-Attainment Planning", 81st Annual Meeting of the Air Pollution Control Assn., 20-24 June 1980, Dallas, TX.

Tesche, T.W., J.G. Wilkinson, D.E. McNally, R. Kapahi, and W.R. Oliver (1988b). "Photochemical Modeling of Two SCCAMP 1984 Oxidant Episodes, II: Modeling Procedures and Evaluation Results," prepared for U.S. Environmental Protection Agency by Radian Corporation, Sacramento, CA.

Tesche, T. W., and M. A. Yocke (1976). "Numerical Modeling of Wind Fields over Mountainous Regions in California", Conference on Sierra Nevada Meteorology, American Meteorological Society, 19-21 June, South Lake Tahoe, CA.